

University of Nebraska - Lincoln

DigitalCommons@University of Nebraska - Lincoln

Cliff Stains Publications

Published Research - Department of Chemistry

2010

Toward a General Approach for RNA-Templated Hierarchical Assembly of Split-Proteins

Jennifer L. Furman
University of Arizona

Ahmed H. Badran
University of Arizona

Oluyomi Ajulo
University of Arizona

Jason R. Porter
University of Arizona

Cliff I. Stains
University of Nebraska-Lincoln, cstains2@unl.edu

See next page for additional authors

Follow this and additional works at: <https://digitalcommons.unl.edu/chemistrystains>

 Part of the [Chemistry Commons](#)

Furman, Jennifer L.; Badran, Ahmed H.; Ajulo, Oluyomi; Porter, Jason R.; Stains, Cliff I.; Segal, David J.; and Ghosh, Indraneel, "Toward a General Approach for RNA-Templated Hierarchical Assembly of Split-Proteins" (2010). *Cliff Stains Publications*. 5.

<https://digitalcommons.unl.edu/chemistrystains/5>

This Article is brought to you for free and open access by the Published Research - Department of Chemistry at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Cliff Stains Publications by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

Authors

Jennifer L. Furman, Ahmed H. Badran, Oluyomi Ajulo, Jason R. Porter, Cliff I. Stains, David J. Segal, and Indraneel Ghosh

Submitted May 20, 2010; published online August 3, 2010.

Toward a General Approach for RNA-Templated Hierarchical Assembly of Split-Proteins

Jennifer L. Furman,¹ Ahmed H. Badran,¹ Oluyomi Ajulo,¹ Jason R. Porter,¹
Cliff I. Stains,¹ David J. Segal,² and Indraneel Ghosh¹

1. Department of Chemistry & Biochemistry, University of Arizona, 1306 East University Boulevard, Tucson, Arizona 85721

2. Genome Center and Department of Pharmacology, University of California, Davis, California 95616

Corresponding author — I. Ghosh, ghosh@email.arizona.edu

Abstract

The ability to conditionally turn on a signal or induce a function in the presence of a user-defined RNA target has potential applications in medicine and synthetic biology. Although sequence-specific pumilio repeat proteins can target a limited set of ssRNA sequences, there are no general methods for targeting ssRNA with designed proteins. As a first step toward RNA recognition, we utilized the RNA binding domain of argonaute, implicated in RNA interference, for specifically targeting generic 2-nucleotide, 3' overhangs of any dsRNA. We tested the reassembly of a split-luciferase enzyme guided by argonaute-mediated recognition of newly generated nucleotide overhangs when ssRNA is targeted by a designed complementary guide sequence. This approach was successful when argonaute was utilized in conjunction with a pumilio repeat and expanded the scope of potential ssRNA targets. However, targeting any desired ssRNA remained elusive as two argonaute domains provided minimal reassembled split-luciferase. We next designed and tested a second hierarchical assembly, wherein ssDNA guides are appended to DNA hairpins that serve as a scaffold for high affinity zinc fingers attached to split-luciferase. In the presence of a ssRNA target containing adjacent sequences complementary to the guides, the hairpins are brought into proximity, allowing for zinc finger binding and concomitant reassembly of the fragmented luciferase. The scope of this new approach was validated by specifically targeting RNA encoding VEGF, hDM2, and HER2. These approaches provide potentially general design paradigms for the conditional reassembly of fragmented proteins in the presence of any desired ssRNA target.

Introduction

The ability to conditionally elicit a response in the presence of a user-defined macromolecular target has great potential in biosensor design, targeted therapeutics, and a variety of applications in synthetic biology. With our enhanced understanding of the human genome, both DNA and RNA have become increasingly important biological targets related to diverse human disease.¹ General methods for nucleic acid targeting now include several hybridization based approaches such as RNA interference (RNAi), triplex-forming oligonucleotides (TFOs), peptide nucleic acids (PNAs), as well as the elegantly designed polyamides.² Nucleic acid detection most commonly relies on fluorescently labeled oligonucleotides, as in the case of fluorescence *in situ* hybridization (FISH).³ However, the constitutive fluorescent signal associated with this class of probes necessitates washing steps and results in reduced sensitivity, which potentially limits the utility of this technique for *in*

vivo imaging.⁴ To address these concerns, recent efforts have been directed toward generating “turn-on” fluorescent or electrochemical sensors, which couple conditional signal output to target binding.⁵ In a complementary approach, there has been recent interest in DNA- and RNA-templated reactions, wherein probe localization on a single-stranded nucleic acid target enables the specific chemical transformation of attached moieties.⁶ Examples of nucleic acid-templated chemical reactions include FRET- and quencher-based autoligation probes, metallosalen–DNA conjugates and deoxyribozymes for DNA hydrolysis, and catalytically released cytotoxic drugs.⁷ Additionally, there is great interest in the design of DNA-directed gain of function proteins, which cannot only serve as sensors and genome modifying agents, but also have potential as specific therapeutics. Many groups have described protein-based approaches for the conditional reassembly of fragmented proteins on double-stranded DNA (dsDNA) mainly utilizing the Cys₂His₂ class of zinc fingers (ZFs) as targeting domains.^{8, 9}

1. (a) Venter, J. C. *Science* 2001, 291, 1304–1351; (b) Sachidanandam, R. *Nature* 2001, 409, 928–933
2. (a) Dervan, P. B. *Biorg. Med. Chem.* 2001, 9, 2215–2235; (b) Hannon, G. J. *Nature* 2002, 418, 244–251; (c) Seidman, M. M. and Glazer, P. M. *J. Clin. Invest.* 2003, 112, 487–494; (d) Nielsen, P. E. *Curr. Opin. Mol. Ther.* 2010, 12, 184–191
3. Levsky, J. M. and Singer, R. H. *J. Cell Sci.* 2003, 116, 2833–2838
4. Bao, G., Rhee, W. J., and Tsourkas, A. *Annu. Rev. Biomed. Eng.* 2009, 11, 25–47

5. (a) Boon, E. M., Ceres, D. M., Drummond, T. G., Hill, M. G., and Barton, J. K. *Nat. Biotechnol.* 2000, 18, 1096–1100; (b) Tan, W., Wang, K., and Drake, T. *Curr. Opin. Chem. Biol.* 2004, 8, 547–553; (c) Chenoweth, D. M., Viger, A., and Dervan, P. B. *J. Am. Chem. Soc.* 2007, 129, 2216–2217; (d) Xiao, Y., Lou, X. H., Uzawa, T., Plakos, K. J. I., Plaxco, K. W., and Soh, H. T. *J. Am. Chem. Soc.* 2009, 131, 15311–15316
6. (a) Li, X. Y. and Liu, D. R. *Angew. Chem., Int. Ed.* 2004, 43, 4848–4870; (b) Silverman, A. P. and Kool, E. T. *Chem. Rev.* 2006, 106, 3775–3789

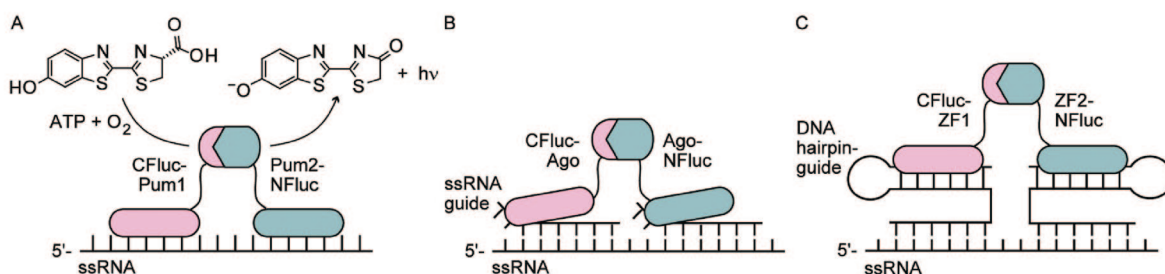


Figure 1. Cartoon representations of pumilio, argonaute, and zinc finger domains attached to split-luciferase. The presence of target ssRNA results in luciferase reassembly and a luminescent signal. (A) Pumilio (Pum) RNA binding proteins confer specificity for a particular RNA sequence. (B) Argonaute (Ago)-mediated detection strategies involve recognition of a 3' 2-nucleotide overhang generated by the binding of short guide sequences of RNA to a complementary target. (C) Zinc fingers (ZF) bind to dsDNA hairpins localized to adjacent sites on an RNA target via hybridization of attached guide sequences. CFluc, C-terminal fragment of luciferase; NFluc, N-terminal fragment of luciferase.

However, the design of general approaches for targeting single-stranded RNA (ssRNA) that elicit the conditional reassembly of functional proteins remains a challenging enterprise. As recently discussed by Varshavsky, the possibility of selectively targeting genetic modifications in malignant cells may be achieved through the implementation of DNA targeted split-protein reassembly based strategies.¹⁰ Herein, we detail our progress toward the long-term goal of RNA dependent conditional split-protein assembly for potentially modulating cell viability by utilizing several iterative designs that combine split-protein methods with nucleic acid targeting and nucleic acid hybridization approaches.

The split-protein reassembly methodology or protein complementation assay (PCA) has been most commonly applied toward the elucidation of protein–protein interactions, wherein conditional protein reassembly is facilitated by the direct interaction of appended domains.¹¹ This methodology, starting with ubiquitin, has been expanded to several monomeric proteins such as the green fluorescent protein (GFP), β -lactamase, luciferases, and tobacco etch virus (TEV) protease.¹² Building on these approaches, we have previously reported a ternary DNA-templated split-protein reassembly system, in which conditional signal output is coupled to the presence of a unique DNA target. In this approach the sequence-specific binding of ZF domains to targeted DNA induces reassembly of the appended split-signaling protein, which comprises GFP and variants, β -lactamase, or firefly luciferase.^{8,13} This strategy has since been adapted for RNA-templated assembly by utilizing sequence-specific RNA binding

proteins called pumilio domains (Figure 1A).^{13b,14} However, there are presently no general approaches for sequence-specifically assembling genetically encoded proteins on any user-defined ssRNA target. Interestingly, in a related DNA-directed approach, split-GFP has been conjugated to single-stranded DNA (ssDNA) oligonucleotides, either chemically or through a biotin/streptavidin linkage, allowing for hybridization-dependent GFP reassembly.¹⁵ In another elegant DNA-directed approach, a split-peroxidase DNAzyme attached to ssDNA domains was designed, which detected 1 μ M complementary DNA.¹⁶ Alternate conditional assembly approaches involve the design of sensitive multicomponent nucleic acid enzymes (MNazymes) or high specificity binary DNA probes.¹⁷ Building upon these studies, we sought to develop a potentially general method that allows for recognizing any ssRNA target, leading to the conditional reassembly of a fragmented protein.

Toward this long-term goal, we herein describe the iterative design of a series of RNA-templated assemblies utilizing split-luciferase, which provides a convenient sensor and surrogate for split-proteins in general.^{13b} In our previous DNA-templated reassembly systems, we achieved target binding through a direct protein–target interaction, thereby inducing split-protein reassembly. To create a more general approach for targeting an expanded set of ssRNA sequences, we initially investigated the use of the PAZ domain of argonaute 2.¹⁸ Argonaute (Ago), which is central to the dicer complex produced during RNAi,² has been reported to bind to the 2-nucleotide (nt), 3' overhangs of short double-stranded RNA (dsRNA).¹⁸ Thus, by providing complementary guide oligonucleotides

7. (a) Ma, Z. C. and Taylor, J. S. *Proc. Natl. Acad. Sci. U.S.A.* 2000, 97, 11159–11163; (b) Czapinski, J. L. and Sheppard, T. L. *J. Am. Chem. Soc.* 2001, 123, 8618–8619; (c) Xu, Y. Z., Karalkar, N. B., and Kool, E. T. *Nat. Biotechnol.* 2001, 19, 148–152; (d) Sando, S., Abe, H., and Kool, E. T. *J. Am. Chem. Soc.* 2004, 126, 1081–1087; (e) Chandra, M., Sachdeva, A., and Silverman, S. K. *Nat. Chem. Biol.* 2009, 5, 718–720; (f) Franzini, R. M. and Kool, E. T. *J. Am. Chem. Soc.* 2009, 131, 16021–16023
8. Stains, C. I., Porter, J. R., Ooi, A. T., Segal, D. J., and Ghosh, I. *J. Am. Chem. Soc.* 2005, 127, 10782–10783
9. (a) Nomura, W. and Barbas, C. F., III. *J. Am. Chem. Soc.* 2007, 129, 8676–8677; (b) Meng, X. D., Noyes, M. B., Zhu, L. H. J., Lawson, N. D., and Wolfe, S. A. *Nat. Biotechnol.* 2008, 26, 695–701
10. Varshavsky, A. *Proc. Natl. Acad. Sci. U.S.A.* 2007, 104, 14935–14940
11. Michnick, S. W., Ear, P. H., Manderson, E. N., Remy, I., and Stefan, E. *Nat. Rev. Drug Discovery* 2007, 6, 569–582
12. (a) Johnsson, N. and Varshavsky, A. *Proc. Natl. Acad. Sci. U.S.A.* 1994, 94, 8405–8410; (b) Ghosh, I., Hamilton, A. D., and Regan, L. *J. Am. Chem. Soc.* 2000, 122, 5658–5659; (c) Galarneau, A., Primeau, M., Trudeau, L. E., and Michnick, S. W. *Nat. Biotechnol.* 2002, 20, 619–622; (d) Paulmurugan, R. and Gambhir, S. S. *Anal. Chem.* 2003, 75, 1584–1589; (e) Luker, K. E., Smith, M. C. P., Luker, G. D., Gammon, S. T., Piwnica-Worms, H., and Piwnica-Worms, D. *Proc. Natl. Acad. Sci. U.S.A.* 2004, 101, 12288–12293; (f) Wehr, M. C., Laage, R., Bolz, U.,

- Fischer, T. M., Grunewald, S., Scheek, S., Bach, A., Nave, K. A., and Rossner, M. *J. Nat. Methods* 2006, 3, 985–993
13. (a) Ooi, A. T., Stains, C. I., Segal, D. J., and Ghosh, I. *Biochemistry* 2006, 45, 3620–3625; (b) Porter, J. R., Stains, C. I., Jester, B. W., and Ghosh, I. *J. Am. Chem. Soc.* 2008, 130, 6488–6497; (c) Furman, J. L., Badran, A. H., Shen, S. Y., Stains, C. I., Hannallah, J., Segal, D. J., and Ghosh, I. *Bioorg. Med. Chem. Lett.* 2009, 19, 3748–3751
14. Ozawa, T., Natori, Y., Sato, M., and Umezawa, Y. *Nat. Methods* 2007, 4, 413–419
15. (a) Demidov, V. V., Dokholyan, N. V., Witte-Hoffman, C., Chalasani, P., Yiu, H.-W., Ding, F., Yu, Y., Cantor, C. R., and Broude, N. E. *Proc. Natl. Acad. Sci. U.S.A.* 2006, 103, 2052–2056; (b) Niemeyer, C. M. *Angew. Chem., Int. Ed.* 2010, 49, 1200–1216
16. (a) Deng, M. G., Zhang, D., Zhou, Y. Y., and Zhou, X. *J. Am. Chem. Soc.* 2008, 130, 13095–13102; (b) Kolpashchikov, D. M. *J. Am. Chem. Soc.* 2008, 130, 2934–2935; (c) Liu, J. W., Cao, Z. H., and Lu, Y. *Chem. Rev.* 2009, 109, 1948–1998
17. (a) Kolpashchikov, D. M. *J. Am. Chem. Soc.* 2005, 127, 12442–12443; (b) Kolpashchikov, D. M. *J. Am. Chem. Soc.* 2006, 128, 10625–10628; (c) Mokany, E., Bone, S. M., Young, P. E., Doan, T. B., and Todd, A. V. *J. Am. Chem. Soc.* 2010, 132, 1051–1059
18. (a) Ma, J.-B., Ye, K., and Patel, D. J. *Nature* 2004, 429, 318–322; (b) Meister, G., Landthaler, M., Patkaniowska, A., Dorsett, Y., Teng, G., and Tuschl, T. *Mol. Cell* 2004, 15, 185–197

similar to those used in RNAi, short regions of dsRNA are generated in the template, creating a platform for Ago binding and split-luciferase reassembly. This quaternary assembly strategy, which utilizes a pumilio domain in conjunction with Ago, provides a basis for potentially recognizing any given RNA target (Figure 1B). In an alternative design, we envisioned that we could attach designed dsDNA hairpins to ssDNA guides complementary to the template, thus, providing a binding site for high affinity ($K_d \sim$ low pM), sequence-specific ZFs appended to split-luciferase (Figure 1C). These "hairpin-guides" would potentially allow ZF-mediated split-luciferase reassembly only in the presence of the ssRNA or ssDNA template. Herein, we demonstrate that these approaches potentially provide a general paradigm for selectively recognizing a variety of user-defined ssRNA targets known to be over-expressed in human disease.

Results and Discussion

RNA-Templated Assembly Using Argonaute. In our quest to expand upon our existing pumilio-based split-firefly luciferase system for RNA-templated assembly (Figure 2A),^{13b} we incorporated an RNA binding protein, the PAZ domain of *Homo sapiens* Ago2. Given that this domain from Ago1 (96% domain similarity) binds with high affinity ($K_d = 2.18$ nM) to the 2-nt, 3' overhangs of short dsRNA,^{18a} we postulated that we would be able to direct our split-proteins toward any RNA template by providing short, complementary guide oligonucleotides similar to those used in RNAi. We chose the RNA sequence bound by the pumilio proteins as our initial nucleic acid target to allow for the direct comparison between designs. This RNA sequence, 5'-CAUGGUGUAUAUAGUCUUUGAUUAUAGC-GGC, comprises two 8-nt pumilio binding sites (underlined), separated by 5-nt and flanked on either side by 5-nt. We initially designed a 16-nt guide (5' guide), which consisted of 14-nt complementary to the 5' end of the RNA template and a 3'UU sequence serving to generate the 2-nt, 3' overhang required for Ago binding (Supporting Information Table S1). A diuridine overhang was selected since this sequence is efficiently recognized by Ago as determined by siRNA gene knockdown experiments.¹⁹ Thus, to test the utility of Ago in this hierarchical assembly strategy, we attached Ago (residues 227–371) through an 18 amino acid linker to the carboxy terminus of the C-terminal luciferase fragment (residues 398–550), generating CLuciferase–Ago. We then produced CLuciferase–Ago and Pum2–NLuciferase in a cell-free translation system, which has been previously used to rapidly express proteins for split-protein approaches.^{13b} We initiated reassembly by hybridizing the 5' guide to the RNA template. Following incubation of the split-luciferase constructs with the prehybridized guide-template complex at 10 nM, a significant signal over background of 18-fold was observed (Figure 2B). To our knowledge, this is the first example of Ago being employed in a hierarchical assembly approach, providing a general building block in nucleic acid-templated reassembly of split-proteins.

In a complementary set of experiments, we generated Ago–NLuciferase by fusing the Ago domain (residues 227–371) through a 17 amino acid linker to the amino terminus of the N-terminal fragment of luciferase (residues 2–416) and designed an appropriate guide sequence (3' guide) for targeting the 3' end of the RNA target, again incorporating a 3'UU overhang for Ago binding (Supporting Information Table S1).

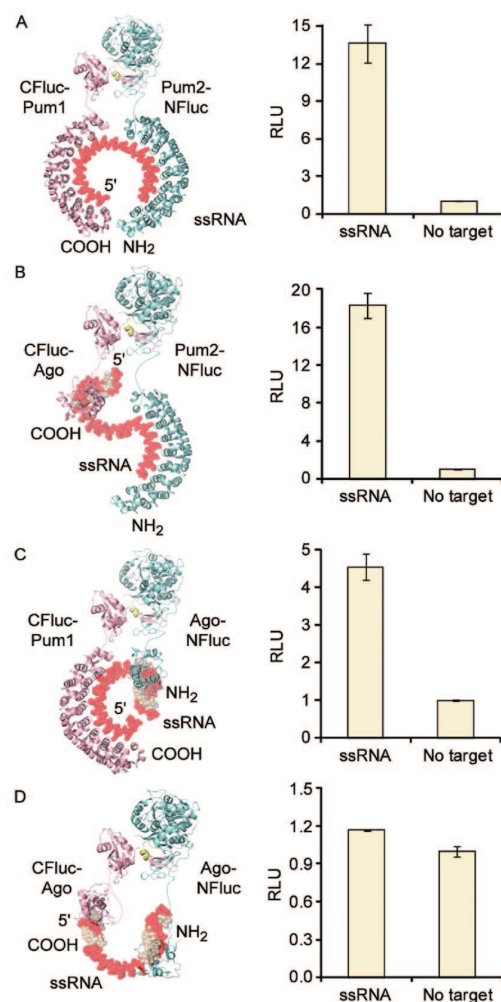


Figure 2. ssRNA-templated reassembly of luciferase using pumilio (Pum) and argonaute (Ago) domains. (A) CLuciferase–Pum1 and Pum2–NLuciferase biosensors were reassembled in the presence of 10 nM cognate RNA. (B) CLuciferase–Ago and Pum2–NLuciferase were reassembled in the presence of 10 nM RNA annealed to the 5' guide. (C) CLuciferase–Pum1 and Ago–NLuciferase were reassembled in the presence of 10 nM RNA annealed to the 3' guide. (D) The RNA target was annealed to the 5' guide and 3' guide to generate proximate Ago binding sites. CLuciferase–Ago and Ago–NLuciferase biosensors were reassembled in the presence of the annealed target (10 nM). In all cases, the CLuciferase (CFluc) fusion proteins are shown in pink, the NLuciferase (NFluc) fusion proteins are shown in blue, the RNA template is red, and the guides are shown in tan. RLU, relative luminescent units.

It has been observed in the crystal structure of *H. sapiens* Ago1 (PDB 1SI2) that, in addition to making significant contacts along the RNA backbone of the strand containing the 3' overhang, Ago also interacts with the 5' terminal residue on the complementary strand.^{18a} In the orientation described for the RNA target and 3' guide, a 5' terminal residue is not generated, which may reduce the affinity of Ago for the nucleic acid complex in this configuration. Despite this potential disadvantage, CLuciferase–Pum1 and Ago–NLuciferase were still successfully used to bind 10 nM target, resulting in luciferase reassembly with a 4.5-fold signal over background (Figure 2C). Finally, to provide a general solution to ssRNA-templated protein assembly exclusively predicated on recognition by Ago domains, we hybridized both guide sequences to the target. However, upon incubation of the ternary guide-template complex at 10 nM with CLuciferase–Ago and Ago–NLu-

19. Elbashir, S. M., Martinez, J., Patkaniowska, A., Lendeckel, W., and Tuschl, T. *EMBO J.* 2001, 20, 6877–6888

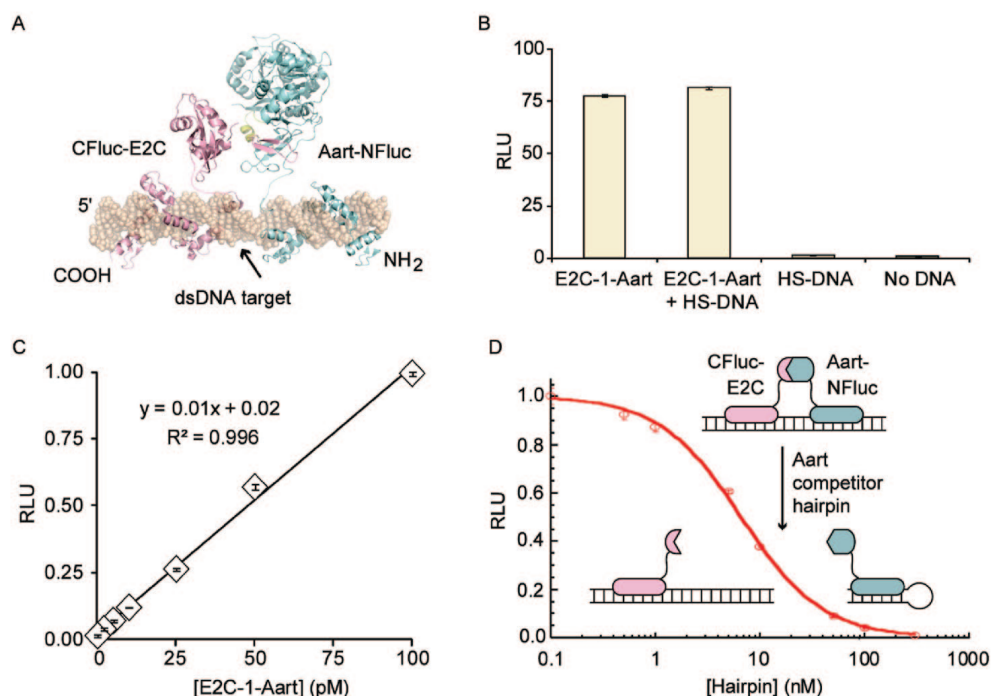


Figure 3. Evaluation of 6-finger zinc fingers E2C and Aart for split-luciferase reassembly. (A) CLuciferase-E2C (pink) and Aart-NLuciferase (blue) bind to a total of 36 base pairs in target DNA (tan), as monitored by a luminescent signal. (B) The CLuciferase-E2C and Aart-NLuciferase biosensors were reassembled in the presence of 300 pM (920 pg) *E2C-1-Aart* DNA, 300 pM *E2C-1-Aart* DNA + 92 ng (100-fold mass excess) herring sperm DNA (HS-DNA), 92 ng HS-DNA only, or No DNA. (C) The CLuciferase-E2C and Aart-NLuciferase biosensors were reassembled in the presence of 100, 50, 25, 10, 5, or 2 pM *E2C-1-Aart* DNA, followed by luminescence readings. The raw data is presented in Supporting Information Table S3. (D) CLuciferase-E2C and Aart-NLuciferase were reassembled in the presence of 300 pM *E2C-1-Aart* DNA, and 300, 100, 50, 10, 5, 1, 0.5, or 0.1 nM Aart competitor hairpin was used to knock down luciferase signal. RLU, relative luminescent units. CFluc, C-terminal fragment of luciferase; NFluc, N-terminal fragment of luciferase.

ciferase, less than a 2-fold signal increase was observed (Figure 2D). We additionally assessed these constructs utilizing a longer 5' guide sequence, as well as larger Ago domains (residues 207–391), with no improvement in signal (Supporting Information Figure S1). The low signal over background appeared to be a function of a reduced luminescent signal in the presence of the ternary nucleic acid template, rather than an increase in background, implying a lack of Ago localization to the intended guides. This may be attributable to the affinity of Ago for a potentially noncognate target as well as the expectation that utilizing both guide sequences in tandem results in a statistical distribution of Ago-luciferase halves bound to the RNA target, permitting only ~50% split-luciferase complementation (Supporting Information Figure S2). Thus, these experiments suggest that Ago may be employed successfully in conjunction with other available sequence-specific RNA binding domains, and its use allows one to access a larger detectable sequence space than pumilio domains alone.

dsDNA-Templated Assembly of High Affinity Zinc Fingers. As a second design, we sought to create a more general, sequence-specific, ssRNA-templated assembly strategy without the need for programming nucleic acid binding proteins for each new ssRNA template. Learning from the Ago-guide strategy, which combines targeted protein and hybridization strategies, we envisioned that attachment of high affinity, sequence-specific ZFs to our split-proteins could serve as a motif for displaying any user-defined ssDNA guide directly appended to a ZF hairpin (HP) binding site, thus, providing a HP-guide. These HP-guides would potentially allow ZF-mediated split-luciferase reassembly only in the presence of a targeted ssRNA or ssDNA sequence. For this pentameric complex to be productively assembled, we reasoned that the interaction between the ZFs and their dsDNA binding sites should be very high affinity to ensure that signal generation is predicated on the interaction between the guides and their

complementary target and is not limited by the interaction of the ZFs with their respective binding sites.

Toward this goal, we investigated the use of two well-characterized and high affinity six-finger ZFs, E2C ($K_d \sim 500$ pM) and Aart ($K_d \sim 7.5$ pM), and initially evaluated these domains for dsDNA-templated assembly (Figure 3A).²⁰ Thus, we utilized a construct in which E2C (residues 7–180) was attached to the carboxy terminus of the C-terminal luciferase fragment (residues 398–550), creating CLuciferase-E2C, while Aart (residues 17–190) was fused to the amino terminus of the N-terminal luciferase fragment (residues 2–416), generating Aart-NLuciferase. We initially tested this system for preferential luciferase reassembly in the presence of an optimized cognate dsDNA template, *E2C-1-Aart* (5'- GCGTAGGGGCCGAGC-CGCAGTGGATGTAGGGAAAAGCCCGGTACCG), consisting of E2C and Aart ZF binding sites (underlined) separated by 1-nt and flanked on either side by 5-nt. Using 300 pM target, a 78-fold signal over background was observed (Figure 3B). Importantly, signal generated from the ternary complex in the presence of a 100-fold mass excess of genomic herring sperm DNA was unperturbed (Figure 3B). Additionally, minimal background signal was generated in the presence of herring sperm DNA alone, suggesting low nonspecific binding for these 6-finger ZFs.

We next performed a series of experiments to characterize the system. dsDNA titrations with decreasing amounts of *E2C-1-Aart* revealed a detection limit of 2 pM (200 amol) that was detectable over the average background signal plus three standard deviations (99% confidence level), thus, providing our most sensitive dsDNA-directed assembly system

20. (a) Beerli, R. R., Segal, D. J., Dreier, B., and Barbas, C. F., III. *Proc. Natl. Acad. Sci. U.S.A.* 1998, 95, 14628–14633; (b) Dreier, B., Beerli, R. R., Segal, D. J., Flippin, J. D., and Barbas, C. F., III. *J. Biol. Chem.* 2001, 276, 29466–29478; (c) Segal, D. J., Crotty, J. W., Bhakta, M. S., Barbas, C. F., III., and Horton, N. C. *J. Mol. Biol.* 2006, 363, 405–421

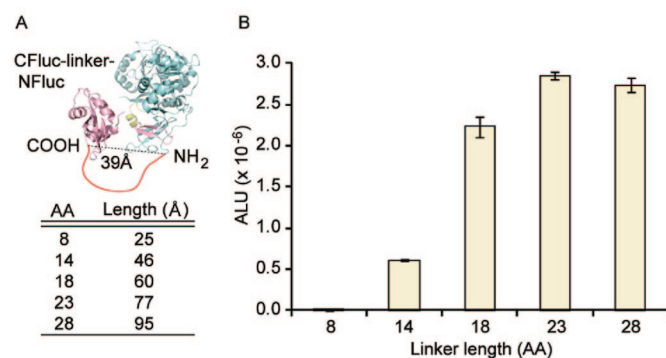


Figure 4. Evaluation of linker length in a covalently tethered luciferase construct. (A) A CLuciferase-linker-NLuciferase construct, comprising residues 398–550 of luciferase (pink) followed by a glycine-serine linker (red) and residues 2–416 of luciferase (blue), was generated to assess the effect of linker length on luciferase reassembly. The distance between the original N- and C-termini of luciferase is ~39 Å, and linkers were designed to vary from ~25 to 95 Å. AA, amino acids. (B) The CLuciferase-linker-NLuciferase constructs were each translated individually in a cell-free system, followed by luminescence readings to determine reassembly efficiency. ALU, arbitrary luminescent units. CFluc, C-terminal fragment of luciferase; NFluc, N-terminal fragment of luciferase.

to date (Figure 3C, Supporting Information Table S3). Finally, we demonstrated the ability of these 6-finger ZFs to bind designed dsDNA HPs, as required for the envisioned RNA-templated pentameric assembly strategy. Translations of the split-proteins, CLuciferase-E2C and Aart-NLuciferase, were initiated in the presence of the target oligonucleotide E2C-1-Aart with or without an Aart competitor HP (Supporting Information Table S1). A competitor HP concentration-dependent decrease in luminescence was observed ($IC_{50} = 7$ nM), confirming that HPs can be designed as ZF binding sites (Figure 3D), thus, providing a platform for the adaptation of this ternary sensor for pentameric RNA-directed assembly.

Investigating Covalently Tethered Luciferase Variants.

To ensure that we had designed an optimal luciferase reassembly system with sufficient flexibility to permit refolding, we systematically evaluated a series of split-luciferase constructs in which the C-terminal fragment of luciferase (residues 398–550) is covalently tethered to the amino terminus of the N-terminal fragment of luciferase (residues 2–416) through a linker of variable length (Figure 4A). Accordingly, the efficiency of luciferase reassembly for this series of CLuciferase-linker-NLuciferase constructs should be predicated on the length of the linker adjoining the inverted halves and will determine approximate linker constraints for future constructs. Considering that the linear distance between the native amino and carboxy termini of luciferase is ~39 Å based on a high-resolution crystal structure (PDB 1LCI), the linker lengths selected were 8 amino acids (AA), 14 AA, 18 AA, 23 AA, and 28 AA (Supporting Information Figure S3), which, respectively, correspond to approximate linear distances (in extended configurations of 3.5 Å per AA) of 25, 46, 60, 77, and 95 Å.²¹ Upon translating each construct in a cell-free system, we observed a maximum luminescent signal at 23 AA, suggesting that 77 Å is sufficient to minimize conformational strain on the system associated with shorter linkers (Figure 4B). A translation performed in the presence of ³⁵S-methionine served to ensure that differences in luminescence were not a function of variable protein expression (Supporting Information Figure

S3). These results imply that the total linker length between each luciferase half and its conjugated domain must at minimum be greater than 23 AA. Our current linkers of 18 and 17 AA for CLuciferase-E2C and Aart-NLuciferase, when considered collectively (35 AA), are in excess of this minimum constraint and as such should provide sufficient length to allow for efficient luciferase reassembly.

ssRNA and ssDNA-Templated Assembly of Split-Luciferase. As noted above, the designed 6-finger ZFs E2C and Aart potentially confer high sensitivity for binding to a pre-defined, cognate dsDNA target. However, despite the availability of ZF motifs that bind ssRNA, programmability of these domains for detection of a user-defined sequence is currently precluded by the lack of a recognition code.²² We therefore sought to harness the sensitivity and specificity of the ZF-dsDNA interaction by utilizing a bridging domain to recognize single-stranded nucleic acid polymers, with an emphasis on ssRNA-templated assembly. As such, we attached ZF HP binding sites to guide sequences of ssDNA, which serve to target adjacent sequences in a user-defined RNA template. Thus, assembly is initiated by the guide sequences binding to proximal sites in their defined target, thereby localizing the ZF binding sites and allowing for binding of the split-luciferase conjugated ZFs (Figure 5A).

We initially tested this pentameric reassembly strategy by targeting the RNA sequence recognized by the pumilio proteins in order to compare results between the different designs. Given the short length of the template (31-nt), we selected guides with lengths of 14- and 13-nt, which have complementary sites in the template separated by 4-nt to prevent any steric conflicts (Figure 5B). The designed HP-guide (E2C-5' guide), which serves as a binding site for E2C and contains a sequence complementary to the 5' end of the RNA target, is 5'-GAGGGGCCG-GAGCCGCGAGTGCCTCGCACTGCGGCTCCGGCCCCCT-CAAACTATATACACCATG. The portions forming the HP dsDNA binding site for E2C (underlined) and connected by a TCCT loop are adjoined to the sequence complementary to the target (bold) by a poly(A)₄ linker. The HP-guide (3' guide-Aart) complementary to the 3' end of the RNA target (bold) and containing a HP binding site for Aart (underlined) is 5'-GCCGC-TATATCAAAAACTCCGGGCTTTTCCTACATGCTCTG-CATGTAGGGAAAAGCCCCGAG. All subsequent HP-guides are derived from these configurations, only varying in the portion complementary to the chosen target (Supporting Information Table S1).

We tested the utility of our pentameric approach in a cell-free translation system to generate our designed, ZF-modified, split-luciferase constructs. Addition of 1 nM ssRNA template hybridized to complementary HP-guides resulted in luciferase reassembly and a signal of 4.5-fold as compared to background (Figure 5C). We additionally demonstrated that the corresponding ssDNA target (1 nM) may also be used to template reassembly with an identical signal over background of 4.5-fold. To directly interrogate sequence selectivity, a single G to T mutation was introduced in the ssDNA, which reduced signal to background levels. Therefore, initial tests indicated that we had successfully developed a pentameric design that allows for the reassembly of functional proteins on specific single-stranded nucleic acid templates.

VEGF RNA-Templated Assembly of Split-Luciferase. To further probe the generality of our ssRNA-templated assembly approach, we selected a physiologically relevant RNA target that comprises nucleotides 109–403 of the vascular endothelial

21. Conti, E., Franks, N. P., and Brick, P. *Structure* 1996, 4, 287–298

22. Hall, T. M. *Curr. Opin. Struct. Biol.* 2005, 15, 367–373

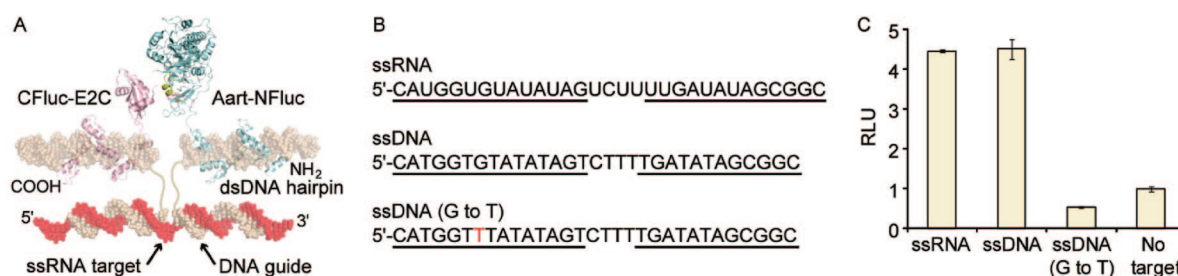


Figure 5. Nucleic acid-templated split-luciferase reassembly. (A) The designed hairpin-guides (tan) target a complementary single-stranded nucleic acid (red), providing proximate binding sites for CLuciferase-E2C (pink) and Aart-NLuciferase (blue) and resulting in generation of a luminescent signal. (B) Target oligonucleotide sequences detected by nucleic acid-templated reassembly. Guides were designed to be complementary to the underlined portions. The single base mutation in the ssDNA target is highlighted in red. (C) Designed hairpin-guides were annealed to 1 nM ssRNA and ssDNA. CLuciferase-E2C and Aart-NLuciferase were incubated with the annealed targets, followed by luminescence readings. RLU, relative luminescent units. CFluc, C-terminal fragment of luciferase; NFluc, N-terminal fragment of luciferase.

growth factor (VEGF), which is overexpressed in certain cancers.²³ Our previous ssRNA target was only 31-nt in length, and we were, therefore, limited in the choice of guide length and binding location. The VEGF transcript generated by *in vitro* transcription was 295-nt in length, and consequently presented a significant expansion of potential guide binding sites. To reduce this collection to a manageable size from which suitable adjacent binding sites could be selected, we initially considered guide length and separation. Keeping in mind that sequence-specificity and binding affinity of probe-target interactions are typically inversely proportional,²⁴ we chose to use tandem guides of 19-nt, which are typically employed in designing siRNAs for use in RNAi.¹⁹ A separation of 2-nt was selected in an attempt to maximize the effective concentration of our localized luciferase halves, while minimizing any steric conflicts. Because probe accessibility can vary greatly within a given target due to the complex nature of RNA secondary structure,²⁵ the final selections for targeting the 295-nt VEGF transcript were made based on the following considerations generally addressed in the design of any complementary oligonucleotide: (1) maximization of thermodynamic stability between guide and target; (2) minimization of inappropriate secondary structures, such as hairpins, loops, cruciforms, triplexes, or quadruplexes in the guide; and (3) minimization of complementarity to other nontarget sites.²⁶

The above constraints led us to identify VEGF positions 60–78 and 81–99 as appropriate target regions. The guide portions of the existing HP-guides that were used to display reassembled luciferase on the previous ssRNA template were modified to incorporate the newly designed ssDNA oligonucleotides complementary to the VEGF transcript (Supporting Information Table S1). Upon translating CLuciferase-E2C and Aart-NLuciferase in our cell-free system, we added our HP-guides annealed to 1 nM *in vitro* transcribed VEGF target and observed a 19-fold signal increase over the presence of HP-guides alone (Figure 6A). To further test the generality of our guide design strategy, we generated a second set of HP-guides for targeting VEGF at positions 216–234 and 237–255

(Supporting Information Table S1). Using these new guides, we similarly achieved signal over background when assayed under identical conditions as the 60–78 and 81–99 set (Supporting Information Figure S4). We next performed a kinetic evaluation of luminescent signal as a function of incubation time with the ternary nucleic acid complex. Interestingly, signal over background was achieved after only 5 min of incubation of the biosensors with the ternary target, providing an extremely rapid assembly method (Figure 6B). From 5 to 30 min, the rate of signal generation increased quickly, followed by only modest gains thereafter. Therefore, all subsequent luminescence readings were collected after a 30 min incubation, where ~10 pM (1 fmol) VEGF RNA was detectable over the average signal plus three standard deviations (99% confidence level) of the HP-guides alone (Figure 6C).

Next, we attempted to optimize our RNA-templated assembly platform by altering the length of the complementary guide sequences in the HP-guides (Supporting Information Table S1). We tested lengths as short as 15-nt for each guide (30-nt total length of complementarity) and as long as 25-nt (50-nt total length of complementarity) (Figure 6D). We identified the original 19-nt guides (38-nt total length of complementarity) as providing the optimal signal over background (Figure 6E). This may be due to a compromise between an improved binding affinity associated with longer guides and a lower propensity for forming undesired secondary structures in shorter guides. Finally, we tested the selectivity of the system by detecting the annealed VEGF RNA target in the presence of a 2000-fold mass excess of *Saccharomyces cerevisiae* (baker's yeast) RNA. Yeast was selected since this organism lacks a complementary binding site (Supporting Information Table S4). The VEGF RNA was selectively detected at 500 pM (5.0 ng) in the presence of 10 µg of yeast RNA, generating a signal (25-fold) as high as in the presence of target alone (Figure 6F, Supporting Information Figure S5). Additionally, the absence of luciferase activity observed in the presence of nonspecific yeast RNA is indicative of the high selectivity of the system afforded by the designed guide sequences.

Detection of hDM2 and HER2 ssRNA Targets. To further establish the generality of our ZF-mediated, ssRNA-templated assembly methodology, we selected two additional ssRNA templates: the human homologue of murine double minute 2 (hDM2) and the human epidermal growth factor receptor 2 (HER2). HP-guides were designed for binding two adjacent 19-nt sequences present in the 549-nt hDM2 and 201-nt HER2 *in vitro* transcribed targets. Under the conditions optimized for VEGF RNA-mediated assembly, the presence of each of the newly introduced targets (1 nM) yielded luminescent signal

23. Plate, K. H., Breier, G., Weich, H. A., and Risau, W. *Nature* 1992, 359, 845–848

24. Demidov, V. V. and Frank-Kamenetskii, M. D. *Trends Biochem. Sci.* 2004, 29, 62–71

25. (a) Fuchs, B. M., Wallner, G., Beisker, W., Schwippl, I., Ludwig, W., and Amann, R. *Appl. Environ. Microbiol.* 1998, 64, 4973–4982; (b) Snyder, T. M., Tse, B. N., and Liu, D. R. *J. Am. Chem. Soc.* 2008, 130, 1392–1401

26. (a) Vickers, T. A., Wyatt, J. R., and Freier, S. M. *Nucleic Acids Res.* 2000, 28, 1340–1347; (b) Martin, S. E. and Caplen, N. J. *Annu. Rev. Genomics Hum. Genet.* 2007, 8, 81–108

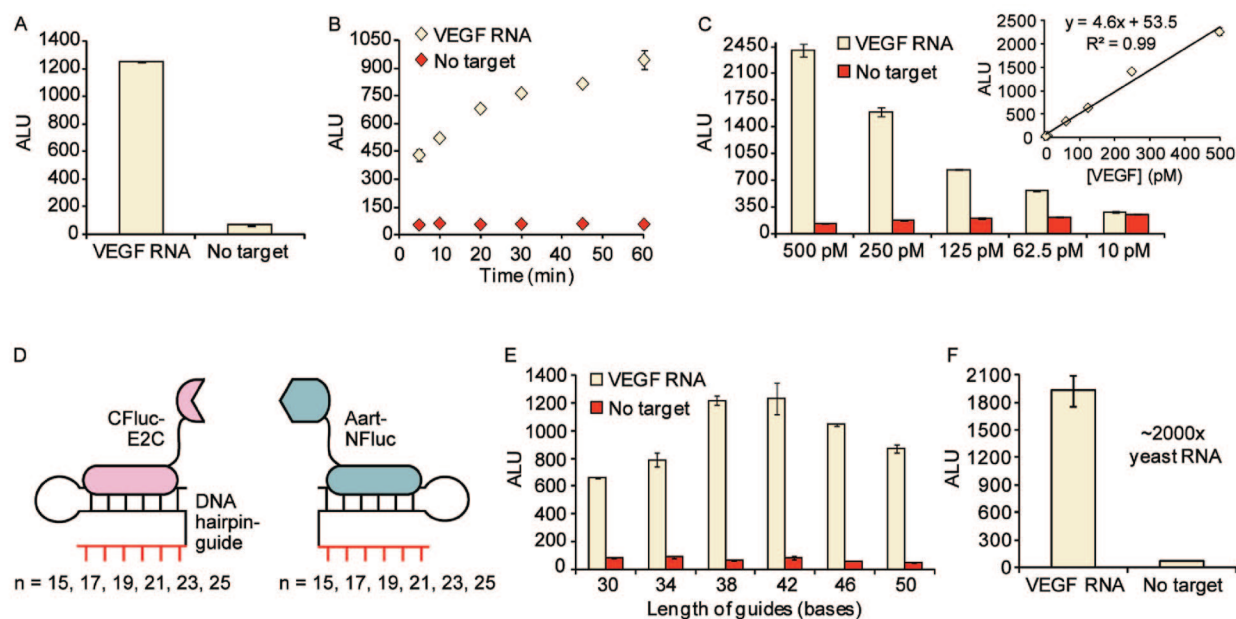


Figure 6. VEGF ssRNA-templated reassembly of split-luciferase. (A) CLuciferase-E2C and Aart-NLuciferase biosensors were reassembled in the presence of a 295-nucleotide (nt) VEGF transcript (1 nM) annealed to designed hairpin-guides with guide lengths of 19-nt. (B) CLuciferase-E2C and Aart-NLuciferase were incubated for 5, 10, 20, 30, 45, or 60 min with the VEGF target (1 nM) annealed to hairpin-guides with guide lengths of 19-nt, followed by luminescence readings. (C) CLuciferase-E2C and Aart-NLuciferase were reassembled in the presence of decreasing concentrations, 500, 250, 125, 62.5, 10 pM, of the VEGF target annealed to hairpin-guides with guide lengths of 19-nt. Upon subtraction of the luminescence contributed by the corresponding No target samples, a linear response to change in RNA concentration is observed (inset). (D) The length of the complementary guides (red) was varied (15, 17, 19, 21, 23, or 25-nt) to identify the optimal length. (E) Pairs of the hairpin-guides described in panel D were annealed to the VEGF target (1 nM), and incubated with CLuciferase-E2C and Aart-NLuciferase, followed by luminescence readings. The length of guides is presented as the combined length of complementarity of a pair of guides. (F) The annealed VEGF target (500 pM = 5.0 ng) was added to ~2000-fold mass excess (10 μ g) of yeast RNA and was incubated in the presence of CLuciferase-E2C and Aart-NLuciferase, followed by luminescence readings. ALU, arbitrary luminescent units. CFLuc, C-terminal fragment of luciferase; NLuc, N-terminal fragment of luciferase.

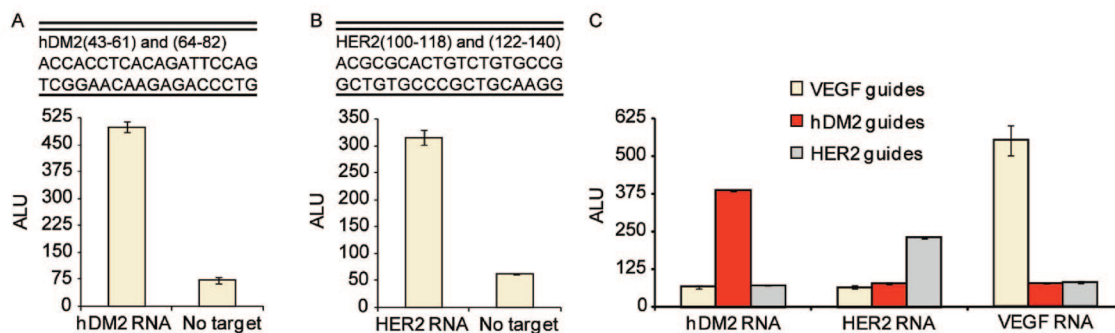


Figure 7. Generality of ssRNA detection. Hairpin-guides were designed to target two additional physiologically relevant transcribed RNAs: (A) human homologue of murine double minute 2 (hDM2) and (B) human epidermal growth factor receptor 2 (HER2). The sequences to which complementary guides were designed are indicated with targeted positions shown in parentheses. The designed hairpin-guides were annealed to their intended targets, followed by incubation of 1 nM annealed target with the CLuciferase-E2C and Aart-NLuciferase biosensors. (C) The hDM2, HER2, and VEGF hairpin-guides were each annealed in the presence of 1 nM hDM2, HER2, and VEGF RNA targets, followed by incubation with CLuciferase-E2C and Aart-NLuciferase. ALU, arbitrary luminescent units.

over the presence of the respective HP-guides alone (Figure 7A,B). These results suggest that, with no other modifications to the reaction conditions, a user may employ any of a number of targets for templated assembly in a straightforward manner by simply selecting an appropriate guide sequence.

With successful reassembly in the presence of each RNA target, we observed slight divergence in the associated sensitivity of binding. Under identical conditions, a 7-fold signal over the presence of HP-guides alone was achieved for the hDM2 target (Figure 7A), and a 5-fold signal was associated with HER2 detection (Figure 7B), as compared to a 19-fold signal for the VEGF template (Figure 6A). Interestingly, these signals do not correlate with target length or predicted melting

temperatures of the guide-target duplex. It is possible that these differences in signal over background that are within 1 order of magnitude may be a function of local secondary structure in the targeted region. In future studies, systematic variation of the targeted region within a given transcript could provide insight into signal deviations and may ultimately provide improved detection limits.

Finally, we tested the orthogonal nature of our RNA-directed assembly strategy. Thus, we attempted annealing of each *in vitro* transcribed target (VEGF, hDM2, or HER2) with each set of HP-guides, both complementary and noncomplementary to the target. Upon incubation with our translated split-proteins, we observed signal only in the cases where the

target was annealed to its complementary set of HP-guides, thus, confirming the high specificity associated with our designed system (Figure 7C).

Conclusion

Conditional enzyme activation under defined cellular conditions has utility in synthetic biology and therapeutic applications. Herein, we have described a series of single-stranded, nucleic acid-templated strategies for conditionally reconstituting split-proteins. Specifically, we have developed a potentially general and robust methodology that utilizes nucleic acid binding domains, including pumilio, argonaute, and zinc fingers, for the templated reassembly of fragmented proteins on user-defined ssRNA and ssDNA targets. Utilizing our previously established pumilio-based split-luciferase reassembly technique as a starting point, we pursued the design of a modular and programmable means for sequence-specifically detecting any single-stranded nucleic acid of interest.^{13b} We initially utilized the RNA binding domain of argonaute, which binds to the 3', 2-nt overhangs of dsRNA targets. By simply designing short guide sequences of RNA complementary to an intended target, an argonaute binding site is generated. This strategy was successful when utilized in conjunction with pumilio domains, but a completely argonaute-based detection system remained elusive.

Therefore, in order to provide a sensitive and programmable solution for ssDNA and ssRNA-templated assembly, a second strategy was employed that coupled the conditional signal generation of reassembled split-luciferase to DNA binding by attached high affinity ZF proteins. The fusion constructs are localized to a single-stranded nucleic acid target through attachment of designed ZF hairpin binding sites to complementary guide sequences. The general nature of this technique was established through the specific detection of VEGF, hDM2, and HER2 mRNA. Notably, this approach allows for distinguishing single base substitutions and detecting the intended target with high specificity, while achieving sensitivities as low as 1 fmol of template. We furthermore anticipate a potential for multiplexing, given the appropriate design of orthogonal assemblies utilizing distinct fragmented proteins, such as different luciferases.²⁷ We also note a potential caveat for the application of this methodology in a cellular context, as this will certainly hinge on the ability to successfully deliver the necessary nucleic acid polymers, though many recent advances have been made in this area.²⁸ An additional consideration must be made regarding potential activation of RNase H in response to formation of the DNA:RNA hybrid produced upon guide binding to the target, which may be potentially avoided by utilizing alternate DNA or RNA targeted nucleic acid analogues.^{2d, 29}

More generally, the utility of nucleic acid-templated reactions lies in a strategy borrowed from nature that allows for controlling the effective concentration of potential reactants. Reactions catalyzed by these techniques range from the classic phosphodiester bond formation to a variety of new and enabling reactions.^{6, 30} Furthermore, the ability to elicit a condi-

tional output in the presence of a nucleic acid target associated with a disease state has far-reaching implications for treating and curing diverse diseases as recently outlined by Varshavsky.¹⁰ Although there have been many elegant strategies that have promise for eliciting a sequence-specific nucleic acid-triggered response in a cellular context, our strategy provides an alternate protein-based platform that has the potential to conditionally activate appropriately fragmented proteins. dsDNA-templating of zinc finger domains has been utilized to reconstitute nucleases, methylases, and integrases, which subsequently modify the genome, as well as localize transcription factors for control of gene expression.^{9, 31} The methodology described herein provides a complementary approach predicated on conditional enzyme activation in the presence of unmodified single-stranded RNA, ultimately providing a mechanism for RNA-templated reassembly of a desired split-protein that results in targeted modulation of cell fate. For example, RNA sequence dependent assembly of enzymes such as β -lactamase, caspase, or RNase has the potential to directly or indirectly affect cell viability.³²

Experimental Methods

Materials. All DNA and RNA oligonucleotides were from Integrated DNA Technologies (IDT) and are listed in the main text or Supporting Information Table S1. RNasin Plus RNase Inhibitor, T7 RiboMAX Large Scale RNA Production kit, Rabbit Reticulocyte Lysate (Promega and Alator), and Steady-Glo Luciferase Assay System were acquired from Promega. Nuclease-free H₂O (Ambion and Promega) was used in all preparations involving RNA.

RNA-Templated Assembly Using Pumilio and Argonaute Domains. The CLuciferase-Pum1 and Pum2-NLuciferase constructs have been previously described.^{13b} The RNA binding PAZ domain (residues 227–371) of *H. sapiens* argonaute 2 (Ago) was cloned into existing plasmids containing split-luciferase (Supporting Information Methods). Genes encoding CLuciferase-Ago and Ago-NLuciferase were PCR amplified, and the corresponding products served as templates for *in vitro* transcription using a T7 Ribomax RNA production kit according to the manufacturer's suggestions. Generally, 3 μ g of amplified DNA template was incubated at 37 °C for 3 h in the presence of 1 \times T7 transcription buffer, 7.5 mM rNTPs, and T7 enzyme mix. To generate argonaute binding sites on the RNA targets, RNA guides, containing regions of complementarity to the pumilio target followed by two 3'-rU's to facilitate Ago binding, were annealed to the pumilio target using the following procedure: heating of target and guide in Buffer A (20 mM Tris-acetate, pH 7.9, 50 mM potassium acetate, 10 mM magnesium acetate, and 1 mM DTT) supplemented with 0.8 U/ μ L RNasin to 90 °C for 1 min, followed by cooling to 37 °C over 1 h. The mRNAs encoding the split-proteins were translated in the Rabbit Reticulocyte Lysate, consisting of 25 μ L reactions prepared in duplicate according to the manufacturer's instructions. A typical reaction was performed at 30 °C for 1.5 h and consisted of the following com-

27. (a) Villalobos, V., Naik, S., and Piwnicka-Worms, D. *Annu. Rev. Biomed. Eng.* 2007, 9, 321–349; (b) Hida, N., Awais, M., Takeuchi, M., Ueno, N., Tashiro, M., Takagi, C., Singh, T., Hayashi, M., Ohmiya, Y., and Ozawa, T. *PLoS One* 2009, 4
28. (a) Sun, Q., Cai, S., and Peterson, B. R. *J. Am. Chem. Soc.* 2008, 130, 10064–10065; (b) Dafik, L., Kalsani, V., Leung, A. K. L., and Kumar, K. *J. Am. Chem. Soc.* 2009, 131, 12091–12093
29. (a) Cerritelli, S. M. and Crouch, R. J. *FEBS J.* 2009, 276, 1494–1505; (b) Bennett, C. F. and Swayze, E. E. *Annu. Rev. Pharmacol. Toxicol.* 2010, 50, 259–293

30. (a) Gartner, Z. J. and Liu, D. R. *J. Am. Chem. Soc.* 2001, 123, 6961–6963; (b) Wrenn, S. J. and Harbury, P. B. *Annu. Rev. Biochem.* 2007, 76, 331–349
31. (a) Ansari, A. Z. and Mapp, A. K. *Curr. Opin. Chem. Biol.* 2002, 6, 765–772; (b) Alwin, S., Gere, M. B., Guhl, E., Effertz, K., Barbas, C. F., Segal, D. J., Weitzman, M. D., and Cathomen, T. *Mol. Ther.* 2005, 12, 610–617; (c) Gordley, R. M., Gersbach, C. A., and Barbas, C. F., III. *Proc. Natl. Acad. Sci. U.S.A.* 2009, 106, 5053–5058
32. (a) Sancho, J. and Fersht, A. R. *J. Mol. Biol.* 1992, 224, 741–747; (b) Chelur, D. S. and Chalfie, M. *Proc. Natl. Acad. Sci. U.S.A.* 2007, 104, 2283–2288

ponents: 66% lysate, 20 μ M each amino acid, 70 mM KCl, 1 mM DTT, 0.8 U/ μ L RNasin, and 0.1–2 pmol each mRNA transcript (Supporting Information Table S2). In the case of CLuciferase–Pum1 and Pum2–NLuciferase, the ssRNA target was added prior to initiation of the translation reaction. In all other cases, the target with annealed guide(s) was added after translation, and binding was allowed to occur for 1 h at 4 °C. Activity was monitored as a luminescent signal produced upon addition of Steady-Glo Luciferase Assay System, where 20 μ L of each translation was added to 80 μ L of Steady-Glo reagent. Luminescence readings were acquired 1 min after mixing using a Turner TD-20e Luminometer with a 10 s integration time. Data is presented as the average of duplicate luminescence readings, and results are representative of at least two independent trials.

dsDNA-Templated Assembly Using 6-Finger Zinc Fingers. Generation of the plasmids encoding CLuciferase–E2C and Aart–NLuciferase is included in the Supporting Information Methods. DNA encoding CLuciferase–E2C and Aart–NLuciferase was amplified from existing plasmids. Transcription was performed using a T7 Ribomax RNA production kit according to the manufacturer's instructions, in which a reaction was incubated at 37 °C for 3 h and typically consisted of 3 μ g of amplified DNA template, 1 \times T7 transcription buffer, 7.5 mM rNTPs, and T7 enzyme mix. To demonstrate DNA-templated reassembly and assess the effect of Herring Sperm DNA (HS-DNA) (Invitrogen), duplicate 25 μ L reactions were carried out in Rabbit Reticulocyte Lysate according to the manufacturer's protocol using 66% lysate, 20 μ M each amino acid, 70 mM KCl, 1 mM DTT, 0.8 U/ μ L RNasin, 10 μ M ZnCl₂, 0.4 pmol of CLuciferase–E2C, 0.2 pmol of Aart–NLuciferase mRNA, and either 300 pM (920 pg) *E2C-1-Aart*, 300 pM *E2C-1-Aart* plus 92 ng HS-DNA, or 92 ng HS-DNA. To determine the limit of DNA detection, duplicate 25 μ L reactions were carried out in Rabbit Reticulocyte Lysate according to the manufacturer's protocol using 66% lysate, 20 μ M each amino acid, 70 mM KCl, 1 mM DTT, 0.8 U/ μ L RNasin, 10 μ M ZnCl₂, 0.1 pmol CLuciferase–E2C mRNA, 0.2 pmol Aart–NLuciferase mRNA, and 100, 50, 25, 10, 5, or 2 pM *E2C-1-Aart* dsDNA. To demonstrate binding to designed dsDNA hairpins, translations included 66% lysate, 20 μ M each amino acid, 70 mM KCl, 1 mM DTT, 0.8 U/ μ L RNasin, 10 μ M ZnCl₂, 0.4 pmol of CLuciferase–E2C, 0.2 pmol of Aart–NLuciferase mRNA, 300 pM *Aart-1-E2C* dsDNA target, and 300, 100, 50, 10, 5, 1, 0.5, or 0.1 nM *Aart* hairpin DNA. In all cases, translations were incubated for 90 min at 30 °C and assayed by adding 80 μ L of Steady-Glo Luciferase Assay System to 20 μ L of translated lysate. Light emission was monitored 1 min after substrate addition using a Turner TD-20e luminometer or a Turner Biosystems 20/20ⁿ luminometer with a 10 s integration time. Data is presented as the average of duplicate samples.

Evaluation of CLuciferase-Linker-NLuciferase Reassembly. The series of CLuciferase-linker-NLuciferase constructs was generated as described in the Supporting Information Methods. DNA encoding each CLuciferase-linker-NLuciferase fusion was PCR amplified, and the corresponding products served as templates for *in vitro* transcription using a T7 Ribomax RNA production kit according to the manufacturer's suggestions. Generally, 3 μ g of amplified DNA template was incubated at 37 °C for 3 h in the presence of 1 \times T7 transcription buffer, 7.5 mM rNTPs, and T7 enzyme mix. The mRNAs encoding the proteins fusions were translated individually in the Rabbit Reticulocyte Lysate, consisting of 25 μ L reactions prepared in duplicate according to the manufacturer's instructions. A typical reaction was performed at 30 °C for

1.5 h and consisted of the following components: 66% lysate, 20 μ M each amino acid, 70 mM KCl, 0.8 U/ μ L RNasin, and 2 pmol CLuciferase-linker-NLuciferase mRNA. Activity was monitored as a luminescent signal produced upon addition of Steady-Glo Luciferase Assay System, where 20 μ L of each translation was added to 80 μ L of Steady-Glo reagent. Luminescence readings were acquired 1 min after mixing using a Turner Biosystems 20/20ⁿ luminometer with a 10 s integration time. Data is presented as the average of duplicate luminescence readings, and results are representative of at least two independent trials.

RNA Target Site Selection. To identify favorable target sites for guide design in an RNA target, we employed the use of multiple analysis programs available free-of-charge on the Internet. A list of potential sites was first generated by analyzing the entire RNA target sequence with the Promega siRNA Target Designer tool (<http://www.promega.com/siRNA Designer/>) and the IDT Antisense Design tool (<http://www.idtdna.com/scitools/applications/antisense/antisense.aspx>). From this collection of sites, pairs were identified that had small (2–3 nt) separations between them. The melting temperature of each oligonucleotide was determined using IDT SciTools OligoAnalyzer 3.1 (<http://www.idtdna.com/analyzer/applications/oligoanalyzer>). Additionally, potential oligonucleotide sequences were analyzed to ensure that they would not anneal to alternate sites in the target using pDRAW32 DNA analysis software (<http://www.acaclone.com>). The entire RNA target was also analyzed by mFold, but long unstructured regions were generally not observed.³³ The information from each analysis source was compiled, and potential sets (generally 5–6) were compared. Ultimately, the set with the highest melting temperature was selected, provided that all other criteria were satisfied.

Generation of Ternary RNA Targets. Single-stranded RNA and DNA oligonucleotide targets were obtained from Integrated DNA Technologies (IDT). DNA encoding VEGF, hDM2, and HER2 was amplified from existing plasmids (Supporting Information Methods). Transcription was performed according to the manufacturer's instructions. Generally, 3 μ g of amplified DNA template was incubated at 37 °C for 3 h in the presence of 1 \times T7 transcription buffer, 7.5 mM rNTPs, and T7 enzyme mix. In all cases, 10 μ M ssRNA target was heat denatured in Buffer B (10 mM Tris-HCl, pH 7.9, 150 mM NaCl, 10 mM MgCl₂, 1 mM DTT) supplemented with 0.8 U/ μ L RNasin at 90 °C for 7 min, followed by cooling at a rate of 6 °C/min to 37 °C for 25 min. Heat denaturation was followed directly by dilution of the target (1 μ M) or blank into equimolar concentrations of the corresponding set of preformed hairpins in Buffer B supplemented with 0.8 U/ μ L RNasin, thus, generating a ternary nucleic acid complex. This mixture was then held at 37 °C for 30 min. DNA hairpins were preformed in Buffer B by an annealing procedure consisting of heating to 95 °C for 7 min, cooling at a rate of 6 °C/min to 56 °C for 5 min, followed by cooling at a rate of 6 °C/min to 25 °C for 10 min. These mixtures were then held at 37 °C for 30 min and stored at 4 °C.

RNA-Templated Assembly of Split-Luciferase. CLuciferase–E2C and Aart–NLuciferase mRNA transcripts were generated as described above. Translations were performed at 30 °C for 1.5 h in the Rabbit Reticulocyte Lysate according to the manufacturer's instructions. A typical reaction consisted of the

33. (a) Mathews, D. H., Sabina, J., Zuker, M., and Turner, D. H. *J. Mol. Biol.* 1999, 288, 911–940; (b) Zuker, M. *Nucleic Acids Res.* 2003, 31, 3406–3415

following components: 66% lysate, 20 μ M each amino acid, 70 mM KCl, 1 mM DTT, 0.8 U/ μ L RNasin, 10 μ M ZnCl₂, 0.2 pmol CLuciferase-E2C mRNA, and 0.2 pmol Aart-NLuciferase mRNA. For the detection of ssRNA and ssDNA, oligonucleotides or VEGF, hDM2, and HER2 RNA annealed nucleic acid ternary complexes (1 nM) or a blank, which contained HP-guides only, were added after completion of the translation reaction. For detection of VEGF in the presence of yeast RNA (Ribonucleic acid mixture from baker's yeast (*S. cerevisiae*) containing mRNA, tRNA, and rRNA, Sigma R6750), translations were performed followed by addition of 500 pM (5.0 ng) annealed VEGF RNA, 500 pM annealed VEGF + 2000-fold mass excess (10 μ g) yeast RNA, yeast RNA + HP-guides, or HP-guides only. To determine a VEGF ssRNA limit of detection, dilutions of the annealed target or HP-guides only (500, 250, 125, 62.5, or 10 pM) were used. In most cases, luminescence readings were taken following 30 min of incubation with the ternary complex at 4 °C. However, to determine a time-depen-

dence of biosensor binding to the ternary VEGF target, incubation times were 5, 10, 20, 30, 45, or 60 min. In all cases, activity was monitored as a luminescent signal produced upon addition of Steady-Glo Luciferase Assay System, where 20 μ L of each translation was added to 80 μ L of Steady-Glo reagent. Luminescence readings were acquired 1 min after mixing using a Turner Biosystems 20/20ⁿ luminometer with a 10 s integration time. Data is presented as the average of duplicate luminescence readings, and results are representative of at least two independent trials.

Acknowledgments — We thank T. Tuschl, D. J. Leahy, and C. F. Barbas, III for the argonaute, HER2, and Aart constructs, respectively. We thank Roy Parker and members of the Ghosh lab for helpful discussions and are grateful to the NIH (GM077403 and CA122630) for financial support. J.L.F and J.R.P were supported by an NIH training grant.

Supporting Information, including supplementary methods, fusion protein sequences, oligonucleotide sequences, prevalence of targeted sites in human and yeast genomes, additional data for reassembly assays, and complete ref 1 citations is presented in the following pages.

Toward a General Approach for RNA-Templated Hierarchical Assembly of Split-Proteins

*Jennifer L. Furman¹, Ahmed H. Badran¹, Oluyomi Ajulo¹, Jason R. Porter¹, Cliff I. Stains¹, David J.
Segal² and Indraneel Ghosh¹**

¹Department of Chemistry, University of Arizona, 1306 East University Blvd, Tucson, AZ 85721

²University of California, Davis, Genome Center and Department of Pharmacology, Davis, CA 95616

* To whom correspondence should be addressed. E-mail: ghosh@email.arizona.edu

Supplementary Information

Supplementary Methods

Materials. Restriction enzymes, dNTPs, antarctic phosphatase, and T4 DNA ligase were purchased from New England Biolabs (NEB). Pfu Ultra polymerase was obtained from Stratagene. All DNA and RNA oligonucleotides were purchased from Integrated DNA Technologies (IDT). RNasin, Steady-Glo® Luciferase Assay System, and Rabbit Reticulocyte Lysate (Promega or Alator) were obtained from Promega.

Cloning. The RNA binding PAZ domain of *Homo sapiens* argonaute-2 (Ago) was PCR amplified from pIRESneo-FLAG/HA Ago2 corrected (Addgene plasmid 10822),¹ which encodes residues 1-856 of *hsAgo2*. Only the RNA binding domain (residues 227-371) was amplified, since adjacent domains have endonuclease activity. This domain of Ago2 is assumed to adopt a stable structure based on the crystal structure of the corresponding domain of Ago1, which is 96% similar.² An additional Ago construct comprising residues 207-391 was also PCR amplified to investigate if a larger domain would be better suited for detection purposes. Ago inserts were ligated to existing plasmids containing CLuciferase and NLuciferase, generating pETDuet-CLuciferase-Ago (Sequence S1) and pETDuet-Ago-NLuciferase (Sequence S2). All methods concerning cloning of Pum2-NLuciferase and CLuciferase-Pum1 have been described.³

To generate the plasmid, pcDNA3.1(+)-CLuciferase-linker-NLuciferase, DNA encoding NLuciferase was PCR amplified from an existing plasmid and inserted into an unmodified pcDNA3.1(+) vector (Invitrogen). Next CLuciferase-linker(18 AA) was amplified from an existing plasmid and was ligated N-terminal to NLuciferase in the pcDNA3.1(+) vector, creating pcDNA3.1(+)-CLuciferase-linker(18 AA)-NLuciferase (Sequence S3). To generate extended linkers, CLuciferase-linker(18 AA) was PCR amplified from the existing construct using a reverse primer that encoded an additional 5 AAs C-terminal to the existing linker. This insert was then introduced in place of the initial CLuciferase-linker(18 AA) DNA, to generate a construct in which the linker was 23 AA. This technique was subsequently repeated using CLuciferase-linker(23 AA)-NLuciferase as a template to generate the 28

AA linker. To generate shorter linkers, complementary oligonucleotides encoding the linker flanked by AgeI and EcoRV restriction sites were annealed and directly ligated to the digested plasmid to replace the 18 AA linker. DNA encoding the zinc finger Aart (residues 17-190) was generated by PCR amplification from an existing plasmid and ligated into a vector containing NLuciferase, generating pETDuet-Aart-NLuciferase (Sequence S4).⁴ DNA encoding the zinc finger E2C (residues 7-180) was generated by PCR amplification from an existing plasmid and ligated into a vector containing CLuciferase, generating pETDuet-CLuciferase-E2C (Sequence S5).⁵ All sequences were confirmed with dideoxynucleotide sequencing at the University of Arizona Sequencing Facility.

In Vitro Transcription. Genes encoding CLuciferase-Ago, Ago-NLuciferase, CLuciferase-E2C, Aart-NLuciferase, and CLuciferase-linker-NLuciferase constructs were PCR amplified using *in vitro* transcription primers containing a T7 RNA Polymerase promoter and a KOZAK sequence in the forward primer and a 3' hairpin loop in the reverse primer. These primers were designed so that the complementary regions had melting temperatures greater than or equal to 70 °C. A typical PCR amplification included an initial heat denaturation of 95 °C for 5 min, followed by 35 cycles of heating to 95 °C, cooling at a rate of 6 °C/min to an annealing temperature of 50 °C. Elongation at 72 °C for 3 minutes completed the cycle. The PCR products were then used as templates for *in vitro* transcription, followed by protein generation in reticulocyte lysate, as described in the main text. Specific mRNA and target conditions used in each of the experiments (CLuciferase-Pum1/Pum2-NLuciferase, CLuciferase-Ago/Pum2-NLuciferase, CLuciferase-Pum1/Ago-NLuciferase, and CLuciferase-Ago/Ago-NLuciferase) are provided in Table S2.

Optimization of the Argonaute Detection Domain. Considering that the Ago domain was only successful as a detection domain when utilized in conjunction with a pumilio domain, we investigated multiple conditions to improve detection. For comparison to the 5' *guide*, we generated 5' *guide*-2, which is complementary to 16-nt at the 5' end of the RNA target (Table S1). This guide was annealed to the RNA target as described in the main text. *In vitro* translations were carried out as described, using 2

pmol each of CLuciferase-Ago and Pum2-NLuciferase or Ago-NLuciferase. The annealed targets were added at 10 nM after translation, followed by luminescence readings as described. The 5' *guide* and 5' *guide-2* provided very similar luminescence readings for both sets of sensors (Figure S1, A and B).

We additionally investigated the use of a larger Ago domain for detection purposes. Specifically, we compared Ago domains consisting of residues 227-371 and residues 207-391. *In vitro* translations were carried out as described, using 0.1 pmol each of CLuciferase-Ago and Pum2-NLuciferase or 2 pmol each of CLuciferase-Ago and Ago-NLuciferase. In the case of CLuciferase-Ago and Pum2-NLuciferase, the annealed target (RNA + 5' *guide*) was added at 10 nM after translation. In the case of CLuciferase-Ago and Ago-NLuciferase, the annealed target (RNA + 5' *guide* + 3' *guide*) was added at 10 nM during the translation. Luminescence readings were acquired as described. The two different Ago constructs provided very similar luminescence readings for both sets of sensors (Figure S1, C and D).

Generation of VEGF, HER2, and hDM2 Target mRNA. VEGF dsDNA was PCR amplified from an existing plasmid, pQE30-VEGF, which contains nucleotides 109-403 of VEGF cDNA, isoform 165. This 295 nucleotide region was amplified using the primers containing a T7 RNA polymerase promoter indicated in Table S1. HER2 dsDNA (nucleotides 785-985, based on Genbank NM_004448) was PCR amplified from an existing plasmid, pSGHV0-HER2,⁶ which contains a portion of the HER2 extracellular domain, using primers indicated in Table S1. Due to the presence of contaminating PCR products, a gel extraction was performed using a QIAquick PCR purification kit (QIAGEN), resulting in isolation of a pure product, as visualized by agarose gel electrophoresis. hDM2 dsDNA nucleotides 306-854 (based on Genbank NM_002392) was amplified from an existing plasmid. *In vitro* transcription for each of these PCR products was carried out as described in the main text.

Sequence S1. DNA encoding CLuciferase-Ago. CLuciferase is in red, the 18 amino acid linker is shown in black, and Ago (residues 227-371) is in green.

```

atgtccggttatgtaaacaatccggaagcgaccaacgccttgattgacaaggatggatgggtacattctggagac
M S G Y V N N P E A T N A L I D K D G W L H S G D
atagcttactgggacgaagacgaacacttcttcatagttgaccgcttgaagtctttaattaaatacaaaggatat
I A Y W D E D E H F F I V D R L K S L I K Y K G Y
caggtggcccccgctgaattggaatcgatatattgttacaacaccccaacatcttcgacgcgggcgtggcaggtctt
Q V A P A E L E S I L L Q H P N I F D A G V A G L
cccgacgatgacgcgggtgaacttcccgccgcgcttgttgttttgaggcacggaaagacgatgacggaaaaagag
P D D D A G E L P A A V V V L E H G K T M T E K E
atcgtggattacgtcgccagtcaggtaacaacccgcgaaaaagttgcgcggaggagttgtgtttgtggacgaagta
I V D Y V A S Q V T T A K K L R G G V V F V D E V
ccgaaaggtcttaccggaaaaactcgacgcaagaaaaatcagagagatcctcataaaggccaagaagggcggaag
P K G L T G K L D A R K I R E I L I K A K K G G K
tccaaattgggcctgcagggcggttcaggcggtgggggttctggcggggggtgggagccccggggcacagccagta
S K L G L Q G G S G G G G S G G G G S P G A Q P V
atcgagtttgtttgtgaagttttggattttaaaagtattgaagaacaacaaaaacctctgacagattcccaaagg
I E F V C E V L D F K S I E E Q Q K P L T D S Q R
gtaaagttttaccaaagaaattaaaggtctaaaggtggagataacgcactgtgggcagatgaagaggaagtaccgt
V K F T K E I K G L K V E I T H C G Q M K R K Y R
gtctgcaatgtgaccgcggcgcccgccagtcaccaaacattcccgcgtgcagcaggagagcgggcagacgggtggag
V C N V T R R P A S H Q T F P L Q Q E S G Q T V E
tgcacgggtggcccgatatttcaaggacaggcacaagttggttctgcgctacccccacctcccatgtttacaagtc
C T V A Q Y F K D R H K L V L R Y P H L P C L Q V
ggacaggagcagaaaacacacctaccttcccctggaggtctgtaacattgtggcaggacaaagatgtattaaaaaa
G Q E Q K H T Y L P L E V C N I V A G Q R C I K K
ttaacggacaatcagacctcaaccatgatcagagcgactgctagggtcg
L T D N Q T S T M I R A T A R S

```

Sequence S2. DNA encoding Ago-NLuciferase. Ago (residues 227-371) is in green, the 17 amino acid

linker is shown in black, and NLuciferase is in red.

```
gcacagccagtaatcgagtttgtttgtgaagttttggattttaaaagtattgaagaacaacaaaaacctctgaca
A Q P V I E F V C E V L D F K S I E E Q Q K P L T
gattcccaaagggtaaagttttaccaaagaaattaaaggtctaaaggtggagataacgcactgtgggcagatgaag
D S Q R V K F T K E I K G L K V E I T H C G Q M K
aggaagtaccgtgtctgcaatgtgacccggcgcccgccagtcaccaaacattcccgctgcagcaggagagcggg
R K Y R V C N V T R R P A S H Q T F P L Q Q E S G
cagacggtggagtgacggtggcccagttttcaaggacaggcacaagttggttctgcgctacccccacctccca
Q T V E C T V A Q Y F K D R H K L V L R Y P H L P
tgtttacaagtcggacaggagcagaaacacacctaccttccctggaggtctgtaacattgtggcaggacaaaga
C L Q V G Q E Q K H T Y L P L E V C N I V A G Q R
tgtattaaaaaattaacggacaatcagacctcaacatgatcagagcgactgctaggtcgaccgggtgggggtggc
C I K K L T D N Q T S T M I R A T A R S T G G G G
ggttcaggcggtgggggttctggtgggggtgggtaccgaagacgccccaaaacataaagaaaggccccggcgccattc
G S G G G G S G G G G T E D A K N I K K G P A P F
tattctctagaggatggaaccgctggagagcaactgcataaggctatgaagagatacgccctgggttcttctggaaca
Y P L E D G T A G E Q L H K A M K R Y A L V P G T
attgcttttacagatgcacatatcgaggtgaacatcacgtacgcggaataacttcgaaatgtccggttcggttggca
I A F T D A H I E V N I T Y A E Y F E M S V R L A
gaagctatgaaacgatatgggctgaatacaaatcacagaatcgctcgatgcagtgaaaactctcttcaattcttt
E A M K R Y G L N T N H R I V V C S E N S L Q F F
atgccggtgttgggcgcttatttatcgaggttgacgttgccgcccgcgaacgcacatttataatgaacgtgaattg
M P V L G A L F I G V A V A P A N D I Y N E R E L
ctcaacagatgaacatttcgcagcctaccgtagtgtttgtttccaaaagggttgcaaaaaattttgaacgtg
L N S M N I S Q P T T V V F V S K K G L Q K I L N V
caaaaaaattaccaataatccagaaaattattatcatggattctaaaacggattaccagggatttcagtcgatg
Q K K L P I I Q K I I I M D S K T D Y Q G F Q S M
tacaggttcgtcacatctcatctacctcccggttttaataacgattttgtaccagagtcctttgatcgtgac
Y T F V T S H L P P G F N E Y D F V P E S F D R D
aaaacaattgcactgataatgaattcctctggatctactgggttacctaagggtgtggcccttccgcatagaact
K T I A L I M N S S G S T G L P K G V A L P H R T
gcctgcgtcagattctcgcatgccagagatcctattttggcaatcaaatacattccggatactgcgattttaagt
A C V R F S H A R D P I F G N Q I I P D T A I L S
gttggtccattccatcacggttttggaatgtttactacactcgcatatgtgatgtggatttcgagtcgtctta
V V P F H H G F G M F T T L G Y L I C G F R V V L
atgtatagatttgaagaagagctgtttttacgatcccttcaggattacaaaattcaaagtgcgttgctagtacca
M Y R F E E E L F L R S L Q D Y K I Q S A L L V P
accctattttcattcttcgccaaaagcactctgattgacaaatacgaatttatctaatttacacgaaattgcttct
T L F S F F A K S T L I D K Y D L S N L H E I A S
gggggcgcacctctttcgaaagaagtcggggaagcgggttgcaaacgcttccatcttccagggatacgacaagga
G G A P L S K E V G E A V A K R F H L P G I R Q G
tatgggctcactgagactacatcagctattctgattacacccgggggatgataaacggggcgcggtcggtaaa
Y G L T E T T S A I L I T P E G D D K P G A V G K
gttggtccattttttgaagcgaaggttggtggatctggataccgggaaaacgctgggcgttaatcagagaggcgaa
V V P F F E A K V V D L D T G K T L G V N Q R G E
ttatgtgtcagaggacctatgattatgtccggttatgtaacaatccggaagcgaccaacgccttgattgacaag
L C V R G P M I M S G Y V N N P E A T N A L I D K
gatgga
D G
```

Sequence S3. CLuciferase-linker(18 AA)-NLuciferase. CLuciferase is shown in red, the 18 amino acid linker is shown in black, and NLuciferase is green. The restriction sites flanking the linker, which were used to facilitate alterations in the linker length, are boxed.

```

atgggtacctccggttatgtaaacaatccggaagcgaccaacgccttgattgacaaggatggatggctacattct
M G T S G Y V N N P E A T N A L I D K D G W L H S
ggagacatagcttactgggacgaagacgaacacttcttcatcggttgaccgcctgaagtctctgattaagtacaaa
G D I A Y W D E D E H F F I V D R L K S L I K Y K
ggctatcaggtggctcccgtgaattggaatccatcttctgctccaacaccccaacatcttcgacgcagggtgtcgca
G Y Q V A P A E L E S I L L Q H P N I F D A G V A
ggtcttcccgcagatgacgcgggtgaacttcccgcgcggttggtgttttgagacacggaagacgatgacggaa
G L P D D D A G E L P A A V V V L E H G K T M T E
aaagagatcggtgattacgtcgccagtcaagtaacaaccgcgaaaaagttgcgcggaggagttgtgtttgtggac
K E I V D Y V A S Q V T T A K K L R G G V V F V D
gaagtaccgaaaggtcttaccgaaaaactcgacgcaagaaaaatcagagagatcctcataaaggccaagaagggc
E V P K G L T G K L D A R K I R E I L I K A K K G
ggaaagatcgccgtgaccggtggcggtgggggttctggcggggggtggggagcggtggcggttctgatatcgaagac
G K I A V T G G G G G S G G G G S G G G S D I E D
gccaaaaacataaagaaaggcccgcgccatttctatccgctggaagatggaaccgctggagagcaactgcataag
A K N I K K G P A P F Y P L E D G T A G E Q L H K
gctatgaagagatacgccctggttccctggaacaattgcttttacagatgcacatatcgaggtggacatcacttac
A M K R Y A L V P G T I A F T D A H I E V D I T Y
gctgagtacttcgaaatgtccgttcggttggcagaagctatgaaacgatatgggctgaatacaaatcacagaatc
A E Y F E M S V R L A E A M K R Y G L N T N H R I
gtcgtatgcagtgaaaactctcttcaattctttatgccggtggtgggcgcgttatttatcgaggttcaggttgcg
V V C S E N S L Q F F M P V L G A L F I G V A V A
cccgcgaacgacatttataatgaacgtgaattgctcaacagtatgggcatttcgcagcctaccgtgggtgttcggt
P A N D I Y N E R E L L N S M G I S Q P T V V F V
tccaaaaaggggttgcaaaaaattttgaacgtgcaaaaaaggtcccaatcatccaaaaattattatcatggat
S K K G L Q K I L N V Q K K L P I I Q K I I I M D
tctaaaacggattaccagggatttcagtcgatgtacacgttcgtcacatctcatctacctcccgggttttaataa
S K T D Y Q G F Q S M Y T F V T S H L P P G F N E
tacgattttgtgccagagtccttcgatagggacaagacaattgcactgatcatgaactcctctggatctactggt
Y D F V P E S F D R D K T I A L I M N S S G S T G
ctgcctaaagggtgctgctctgcctcatagaactgcctgctgagattctcgcatgccagagatcctatttttggc
L P K G V A L P H R T A C V R F S H A R D P I F G
aatcaaatcattccggatactgcgatttttaagtgttgttccattccatcacgggttttggaatgtttactacactc
N Q I I P D T A I L S V V P F H H G F G M F T T L
ggatatttgatatgtggatttcgagtcgtcttaattgtatagatttgaaagaagagctgtttctgaggagccttcag
G Y L I C G F R V V L M Y R F E E E L F L R S L Q
gattacaagattcaaagtgcgctgctgggtgccaaccctattctccttcttcgcaaaaagcactctgattgacaaa
D Y K I Q S A L L V P T L F S F F A K S T L I D K
tacgattttatctaatttacacgaaattgcttctggtggcgctcccctctctaaggaagtcgggggaagcggttgc
Y D L S N L H E I A S G G A P L S K E V G E A V A
aagaggttccatctgccaggtatcaggcaaggatatgggctcactgagactacatcagctattctgattacaccc
K R F H L P G I R Q G Y G L T E T T S A I L I T P
gagggggatgataaacggggcgcggtcggttaaagttgttccattttttgaagcgaaggttgtggatctggatacc
E G D D K P G A V G K V V P F F E A K V V D L D T
gggaaaacgctgggcgttaataaagaggcgaaactgtgtgtgagaggtcctatgattatgtccggttatgtaaac
G K T L G V N Q R G E L C V R G P M I M S G Y V N
aatccggaagcgaccaacgccttgattgacaaggatgga
N P E A T N A L I D K D G

```

Sequence S4. DNA encoding Aart-NLuciferase. Aart is in green, the 17 amino acid linker is in black, and NLuciferase is shown in red.

```

ccccgggagagaagccctatgcttgctcgggaatgtggtaagtccttcagccgcagcgatcacctggccgaacaccag
P G E K P Y A C P E C G K S F S R S D H L A E H Q
cgtacccacacgggtgaaaaaccggtataaatgccagagtgccggcaaattcttttagcgataagaaagatctgacc
R T H T G E K P Y K C P E C G K S F S D K K D L T
cggcatcaacgcactcatactggcgagaagccatacaaatgtccagaatgtggcaagtctttcagccagcgcgca
R H Q R T H T G E K P Y K C P E C G K S F S Q R A
aacctgcgcgcccaccaacgtactcacaccggggagaagcccttatgcttgctcgggaatgtggtaagtccttctct
N L R A H Q R T H T G E K P Y A C P E C G K S F S
cagctggccacactgcgcgcccaccagcgtagccacacgggtgaaaaaccggtataaatgccagagtgccggcaaa
Q L A H L R A H Q R T H T G E K P Y K C P E C G K
tcttttagccgcgaggataacctgcacacccatcaacgtactcatactggcgagaagccatacaaatgtccagaa
S F S R E D N L H T H Q R T H T G E K P Y K C P E
tgtggcaagtcttttctccgcgcgcatgctctgaacgtgcaccaacgtactcacaccggcaaaaaaactagcacc
C G K S F S R R D A L N V H Q R T H T G K K T S T
gggtgggggtggcggttcaggcggtgggggttctggtgggggtgggtaccgaagacgcgcaaaaacataaagaaagggc
G G G G G S G G G G S G G G G T E D A K N I K K G
ccggcgccattctatcctctagaggatggaaccgctggagagcaactgcataaggctatgaagagatacgccctg
P A P F Y P L E D G T A G E Q L H K A M K R Y A L
gttcttgaacaattgcttttacagatgcacatatcgaggtgaacatcacgtacgcggaataacttcgaaatgtcc
V P G T I A F T D A H I E V N I T Y A E Y F E M S
gttcggttggcagaagctatgaaacgatatgggctgaatacaaatcacagaatcgctcgatgcagtgaaaactct
V R L A E A M K R Y G L N T N H R I V V C S E N S
cttcaattctttatgcgggtgttggcgcggttatttatcgaggtgcagttgcgcccgcgaacgcacattataat
L Q F F M P V L G A L F I G V A V A P A N D I Y N
gaacgtgaattgctcaacagtatgaacatttcgcagcctaccgtagtgtttggtttccaaaaaggggttgcaaaaa
E R E L L N S M N I S Q P T V V F V S K K G L Q K
attttgaaacgtgcaaaaaaaattaccaataatccagaaaaattattatcatggattctaaaacggattaccagggg
I L N V Q K K L P I I Q K I I I M D S K T D Y Q G
tttcagtcgatgtacacgttcgctcacatctcatctacctcccggttttaataatgaatacgattttgtaccagagtcc
F Q S M Y T F V T S H L P P G F N E Y D F V P E S
tttgatcggtgacaaaacaattgcactgataatgaattcctctggatctactgggttacctaaggggtgtggccctt
F D R D K T I A L I M N S S G S T G L P K G V A L
ccgcatagaactgcctgcgtcagattctcgcatgccagagatcctatttttggcaatcaaatcattccggataact
P H R T A C V R F S H A R D P I F G N Q I I P D T
gcgatttttaagtgttggttccattccatcacggtttttggaatgtttactacactcggtatatttgatgtgtgattt
A I L S V V P F H H G F G M F T T L G Y L I C G F
cgagtcgctttaatgtatagatttgaagaagagctgtttttacgatcccttcaggattacaaaattcaaagtgcg
R V V L M Y R F E E E L F L R S L Q D Y K I Q S A
ttgctagtaccaaccctattttcattcttcgcaaaaagcactctgattgacaaatacgatttatctaattttacac
L L V P T L F S F F A K S T L I D K Y D L S N L H
gaaattgcttctgggggcgcacctctttcgaaagaagtcgggggaagcggttgcaaaacgcttccatcttccaggg
E I A S G G A P L S K E V G E A V A K R F H L P G
atacgacaaggatattgggtcactgagactacatcagctattctgattacacccgagggggatgataaacggggc
I R Q G Y G L T E T T S A I L I T P E G D D K P G
gcggctcggttaaagtgttccattttttgaagcgaaggttgtggatctggataccggggaaaacgctggcggttaat
A V G K V V P F F E A K V V D L D T G K T L G V N
cagagaggcgaattatgtgtcagaggacctatgattatgtccggttatgtaaacaatccggaagcgaccaacgcc
Q R G E L C V R G P M I M S G Y V N N P E A T N A
ttgattgacaaggatgga
L I D K D G

```


Sequence S5. DNA encoding CLuciferase-E2C. CLuciferase is in red, the 14 amino acid linker is in black, and E2C is shown in green.

```
atgtccggttatgtaaacaatccggaagcgaccaacgccttgattgacaaggatggatggctacattctggagac
M S G Y V N N P E A T N A L I D K D G W L H S G D
atagcttactgggacgaagacgaacacttcttcatagttgaccgcttgaagtctttaattaaatacaaaggatat
I A Y W D E D E H F F I V D R L K S L I K Y K G Y
caggtggcccccgctgaattggaatcgatattgtttacaacaccccaacatcttcgacgcgggcgtggcaggtctt
Q V A P A E L E S I L L Q H P N I F D A G V A G L
cccgcgatgacgcgggtgaacttcccgccgcgcttgttgttttggagcacggaaagacgatgacggaaaaagag
P D D D A G E L P A A V V V L E H G K T M T E K E
atcgtggattacgtcgccagtcaagtaacaacccgcgaaaaagttgcgcggaggagttgtgtttgtggacgaagta
I V D Y V A S Q V T T A K K L R G G V V F V D E V
ccgaaaggtcttaccggaaaaactcgacgcgaagaaaaatcagagagatcctcataaaggccaagaagggcggaag
P K G L T G K L D A R K I R E I L I K A K K G G K
tccaaattgggcctgcagggcggttcaggcggtgggggttctggcggggggtgggagcccggggagaagccctat
S K L G L Q G G S G G G G S G G G G S P G E K P Y
gcttgtccggaatgtggttaagtccttcagtaggaaggattcgccttgtgaggcaccagcgtacccacacgggtgaa
A C P E C G K S F S R K D S L V R H Q R T H T G E
aaaccgtataaatgccagagtgcggcaaatcttttagtcagtcgggggatcttaggcgtcatcaacgcactcat
K P Y K C P E C G K S F S Q S G D L R R H Q R T H
actggcgagaagccatacaaatgtccagaatgtggcaagtctttcagtgattgtcgtgatcttgcgaggcaccaa
T G E K P Y K C P E C G K S F S D C R D L A R H Q
cgtactcacaccgggggagaagccctatgcttgtccggaatgtggtaagtccttctctcagagctctcacctggtg
R T H T G E K P Y A C P E C G K S F S Q S S H L V
cgccaccagcgtacccacacgggtgaaaaaccgtataaatgccagagtgcgggcaaatcttttagtgactgccgc
R H Q R T H T G E K P Y K C P E C G K S F S D C R
gaccttgctcgccatcaacgcactcatactggcgagaagccatacaaatgtccagaatgtggcaagtctttcagc
D L A R H Q R T H T G E K P Y K C P E C G K S F S
cgctctgacaagctggtgcgtcaccaacgtactcacaccggtaaaaaaaactagt
R S D K L V R H Q R T H T G K K T S
```

Supplementary Tables

RNA guides	
5' guide	5'CUAUUAUACACCAUGUU
3' guide	5'GCCGCUAUUAUCAAUU
5' guide-2	5'GACUAUUAUACACCAUGUU
Aart hairpin	
	5'GCATGTAGGGAAAAGCCCCGCGTCCTCGCCGGGCTTTTCCCTACATGC
RNA target primers	
VEGF FWD	5'GCAGCTAATACGACTCACTATAGGCATCACGAAGTGGTGAAGTTCATGGATGTCTATCAGC
VEGF REV	5'CTTTCTTTGGTCTGCATTACATTTGTTGTGTGTAGGAAGC
HER2 FWD	5'GCAGCTAATACGACTCACTATAGGCTGATAGACACCAACCGCTCTCGGGC
HER2 REV	5'GTGCTTGGGGCCCCGTGCAGC
hDM2 FWD	5'GCAGCTAATACGACTCACTATAGGATGTGCAATACCAACATGTCTGTACCTACTGATGGTG
hDM2 REV	5'GCGTTTTCTTTGTCGTTACCAGATAATTCATCTGAATTTTCTTCTGTCTCAC
Hairpin-guides	
VEGF(64-78)-E2C	5'GAGGGGCCGGAGCCGCAGTGCGTCCTCGCACTGCGGCTCCGGCCCCCTCAAAAGAAGATGTCCACCAG
VEGF(81-95)-Aart	5'TCATCAGGGTACTCCAAAACCTCCGGGCTTTTCCCTACATGCTCCTGCATGTAGGGAAAAGCCCCGAG
VEGF(62-78)-E2C	5'GAGGGGCCGGAGCCGCAGTGCGTCCTCGCACTGCGGCTCCGGCCCCCTCAAAAGAAGATGTCCACCAGGG
VEGF(81-97)-Aart	5'TCTCATCAGGGTACTCCAAAACCTCCGGGCTTTTCCCTACATGCTCCTGCATGTAGGGAAAAGCCCCGAG
VEGF(60-78)-E2C	5'GAGGGGCCGGAGCCGCAGTGCGTCCTCGCACTGCGGCTCCGGCCCCCTCAAAAGAAGATGTCCACCAGGGTCTC
VEGF(81-99)-Aart	5'GATCTCATCAGGGTACTCCAAAACCTCCGGGCTTTTCCCTACATGCTCCTGCATGTAGGGAAAAGCCCCGAG
VEGF(58-78)-E2C	5'GAGGGGCCGGAGCCGCAGTGCGTCCTCGCACTGCGGCTCCGGCCCCCTCAAAAGAAGATGTCCACCAGGGTCTC
VEGF(81-101)-Aart	5'TCGATCTCATCAGGGTACTCCAAAACCTCCGGGCTTTTCCCTACATGCTCCTGCATGTAGGGAAAAGCCCCGAG
VEGF(56-78)-E2C	5'GAGGGGCCGGAGCCGCAGTGCGTCCTCGCACTGCGGCTCCGGCCCCCTCAAAAGAAGATGTCCACCAGGGTCTCGA
VEGF(81-103)-Aart	5'ACTCGATCTCATCAGGGTACTCCAAAACCTCCGGGCTTTTCCCTACATGCTCCTGCATGTAGGGAAAAGCCCCGAG
VEGF(54-78)-E2C	5'GAGGGGCCGGAGCCGCAGTGCGTCCTCGCACTGCGGCTCCGGCCCCCTCAAAAGAAGATGTCCACCAGGGTCTCGATT
VEGF(81-105)-Aart	5'GTACTCGATCTCATCAGGGTACTCCAAAACCTCCGGGCTTTTCCCTACATGCTCCTGCATGTAGGGAAAAGCCCCGAG
VEGF(216-243)-E2C	5'GAGGGGCCGGAGCCGCAGTGCGTCCTCGCACTGCGGCTCCGGCCCCCTCAAAAGCCTTGGTGAGGTTTGATC
VEGF(237-255)-Aart	5'GCTCATCTCTCTATGTGCAAAAACCTCCGGGCTTTTCCCTACATGCTCCTGCATGTAGGGAAAAGCCCCGAG
HER2(100-118)-E2C	5'GAGGGGCCGGAGCCGCAGTGCGTCCTCGCACTGCGGCTCCGGCCCCCTCAAAACGGCACAGACAGTGC GCGT
HER2(122-140)-Aart	5'CCCTTGACGCGGGCACAGCAAAAACCTCCGGGCTTTTCCCTACATGCTCCTGCATGTAGGGAAAAGCCCCGAG
hDM2(43-61)-E2C	5'GAGGGGCCGGAGCCGCAGTGCGTCCTCGCACTGCGGCTCCGGCCCCCTCAAAACTGGAATCTGTGAGGTGGT
hDM2(64-82)-Aart	5'CCAGGGTCTCTTGTTCGCAAAAACCTCCGGGCTTTTCCCTACATGCTCCTGCATGTAGGGAAAAGCCCCGAG

Table S1. Oligonucleotides used in this study. For the Aart hairpin, the complementary regions forming the Aart binding site are shown in red, and the loop is in green. For the RNA target primers, the T7 RNA Polymerase promoter is shown in blue. For the hairpin-guides, the complementary hairpin sequences are shown in red, the loop is in green, and the guide sequence is in blue. The numbers in parentheses

indicate the intended target binding location relative to the beginning of the transcript, which is designated as position 1. Unless otherwise specified, all VEGF experiments use the VEGF(60-78)-E2C and VEGF(81-99)-Aart guides.

Table S2. *In vitro* translation conditions used corresponding to the data presented in Figures 2A-D of the main text.

mRNA 1	mRNA 2	Final target concentration
2 pmol Pum2-NLuciferase	2 pmol CLuciferase-Pum1	10 nM target
0.1 pmol Pum2-NLuciferase	0.1 pmol CLuciferase-Ago	10 nM 5' guide + 10 nM target
1 pmol Ago-NLuciferase	1 pmol CLuciferase-Pum1	10 nM 3' guide + 10 nM target
2 pmol Ago-NLuciferase	2 pmol CLuciferase-Ago	10 nM 5' guide + 10 nM 3' guide + 10 nM target

Table S3. Data for the E2C-1-Aart detection limit corresponding to Figure 3C of the main text.

[E2C-1-Aart] (pM)	sample 1		sample 2	
	reading 1	reading 2	reading 1	reading 2
100	153	159	155.5	152.2
50	88	86	87.1	93.7
25	41	44	38.7	41.3
10	18	19	18.7	18.7
5	11	12	9.9	10.7
2	6.1	7.1	5.9	5.2
0	2.5	2.8	2.1	1.6

Table S4. Presence of guide sites in yeast and human genomes. Sites shown in blue were used in a custom algorithm (b-site, Segal and Korf, unpublished) to search for sites with up to two mismatches in the genomes of *Saccharomyces cerevisiae* (Build 2.1) and human (Build 37.1). The number of sites is shown with numbers of mismatches in parentheses.

Hairpin-guides		Sites in <i>Saccharomyces cerevisiae</i>	Sites in <i>Homo sapiens</i>
VEGF(60-78)-E2C	5'GAAGATGTCCACCAGGGTC	0	1 (0) 2 (1) 37 (2)
VEGF(81-99)-Aart	5'GATCTCATCAGGGTACTCC	0	1 (0) 4 (1) 70 (2)
VEGF(216-243)-E2C	5'GCCTTGGTGAGGTTTGATC	0	1 (0) 2 (1) 23 (2)
VEGF(237-255)-Aart	5'GCTCATCTCTCCTATGTGC	0	2 (0) 1 (1) 36 (2)
HER2(100-118)-E2C	5'CGGCACAGACAGTGCGCGT	0	1 (0) 0 (1) 2 (2)
HER2(122-140)-Aart	5'CCCTTGCAGCGGGCACAGC	0	3 (0) 0 (1) 41 (2)
hDM2(43-61)-E2C	5'CTGGAATCTGTGAGGTGGT	0	1 (0) 4 (1) 145 (2)
hDM2(64-82)-Aart	5'CCAGGGTCTCTTGTTCGA	0	1 (0) 0 (1) 47 (2)
GAL4 (<i>Sc</i> control seq)	5'TGTTTAAACTTAAGGGATC	1 (0)	N/D

Supplementary Figures

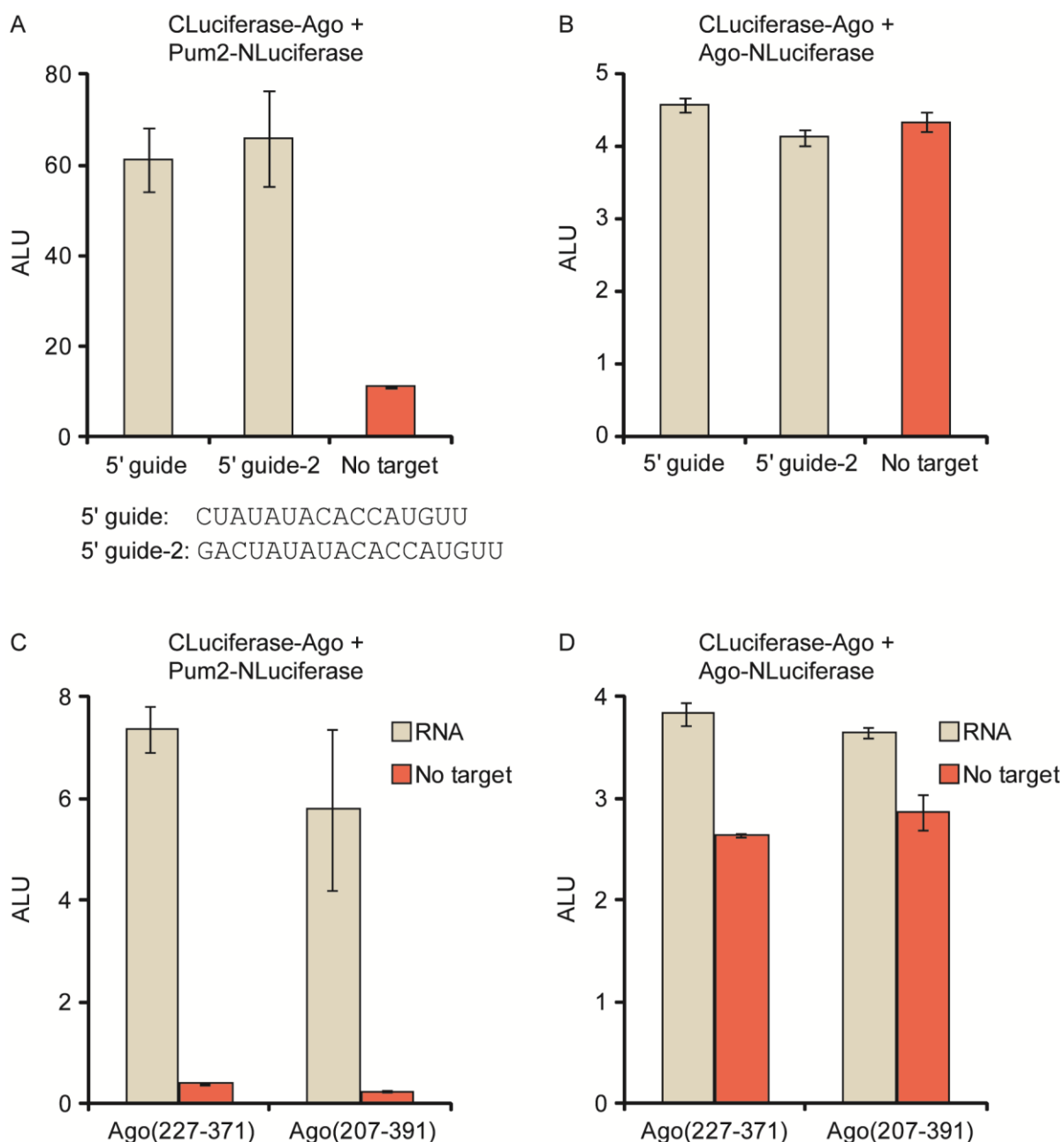


Figure S1. Optimization of the argonaute detection domain. (A) CLuciferase-Ago (Ago residues 227-371) and Pum2-NLuciferase were reassembled in the presence of the RNA target hybridized to either 5' *guide* or 5' *guide-2*, followed by luminescence readings (ALU). (B) CLuciferase-Ago and Ago-NLuciferase were reassembled in the presence of the RNA target hybridized to 3' *guide* and either 5' *guide* or 5' *guide-2*, followed by luminescence readings. (C) Two different Ago constructs, residues 207-391 or 227-371, were utilized as detection domains. CLuciferase-Ago(207-391) or CLuciferase-Ago(227-371) and Pum2-NLuciferase were reassembled in the presence of the RNA target hybridized to the 5' *guide*, followed by luminescence readings. (D) CLuciferase-Ago(207-391) and Ago(207-391)-NLuciferase or CLuciferase-Ago(227-371) and Ago(227-371)-NLuciferase were reassembled in the presence of the RNA target hybridized to the 3' *guide* and 5' *guide*, followed by luminescence readings.

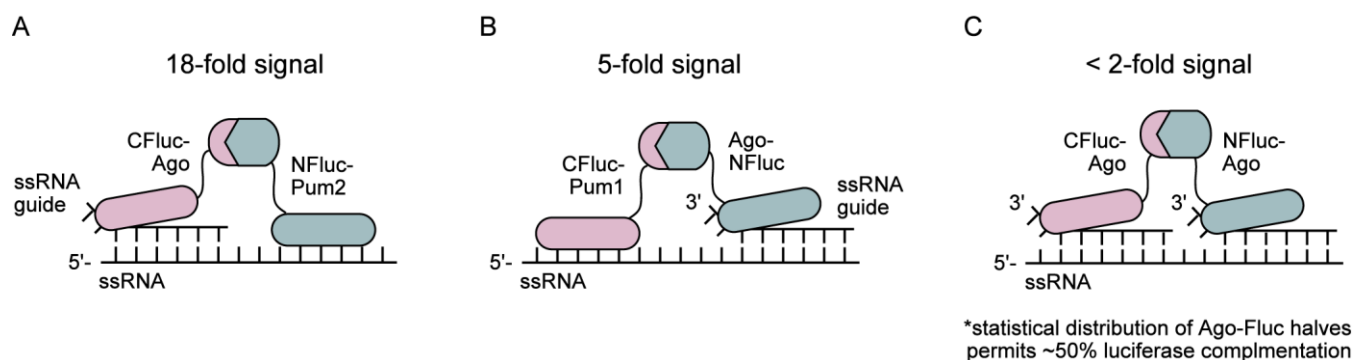


Figure S2. Limitations to RNA-templated split-luciferase reassembly using the Ago domain. (A) Ago can function effectively (18-fold signal) as a detection domain when a guide sequence is provided that is complementary to the 5' end of an RNA target. (B) When the 3' end of the guide sequence is present at an internal site in the target RNA, Ago affinity is reduced as judged by a reduced luminescence signal (4.5-fold signal). (C) Utilizing both guide sequences in tandem results in the expectation that a statistical distribution of Ago-luciferase halves bound to the RNA target would only permit ~50% split-luciferase complementation.

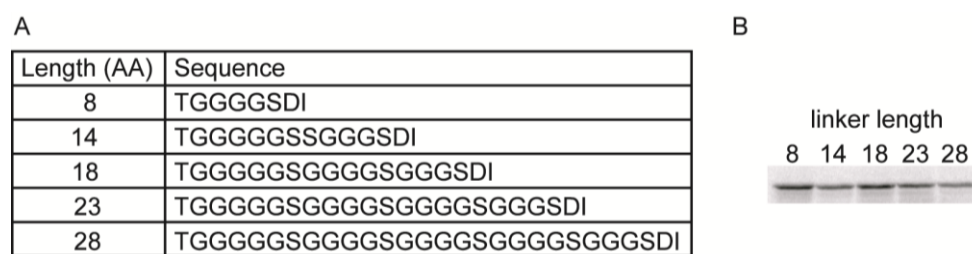


Figure S3. CLuciferase-linker-NLuciferase constructs. (A) The amino acid (AA) sequences of the linkers between CLuciferase and NLuciferase in the pcDNA-3.1(+)-CLuciferase-linker-NLuciferase constructs are presented. (B) Translations of CLuciferase-linker-NLuciferase constructs with 8, 14, 18, 23, and 28 AA linkers were performed in the presence of ^{35}S -methionine, and results were analyzed by SDS-PAGE.

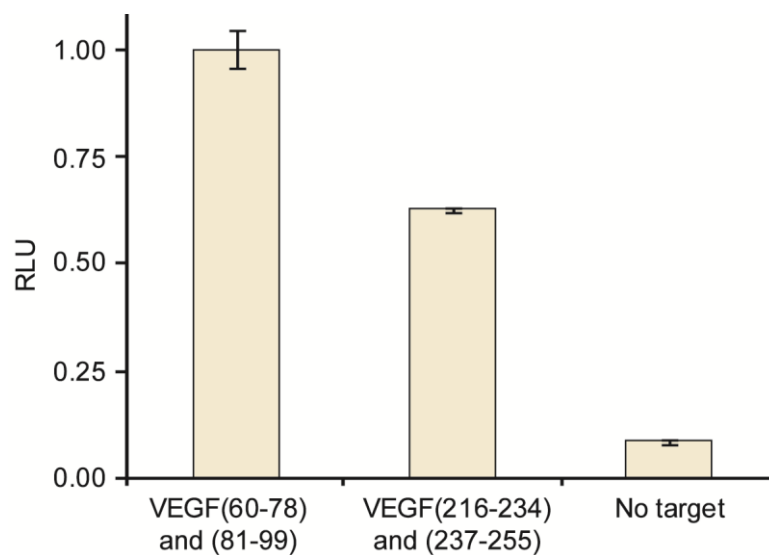


Figure S4. Generality of HP-guide design. HP-guides targeting regions 60-78/81-99 or 216-234/237-255 in the VEGF RNA template were annealed in the presence of the VEGF target. CLuciferase-E2C and Aart-NLuciferase sensors were incubated with the annealed products (1 nM), followed by luminescence readings. Results are presented as luminescence readings relative to the maximal signal (RLU).

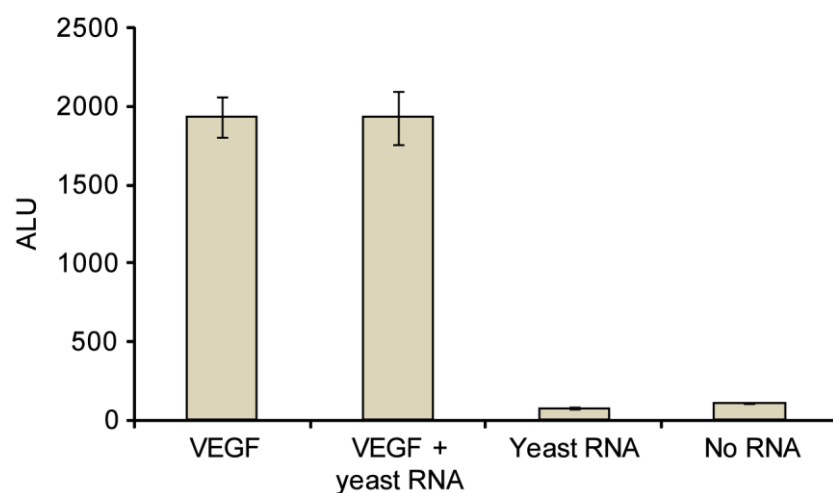


Figure S5. VEGF detection in the presence of yeast RNA. CLuciferase-E2C and Aart-NLuciferase were reassembled in the presence of 500 pM (5.0 ng) annealed VEGF target, 500 pM annealed VEGF + 10 μ g yeast RNA (2000-fold mass excess), 10 μ g yeast RNA, or No RNA. Luminescence readings (ALU) revealed that yeast RNA does not significantly affect VEGF RNA detection. The data presented in Figure 6F of the main text is a sub-set of the data presented here.

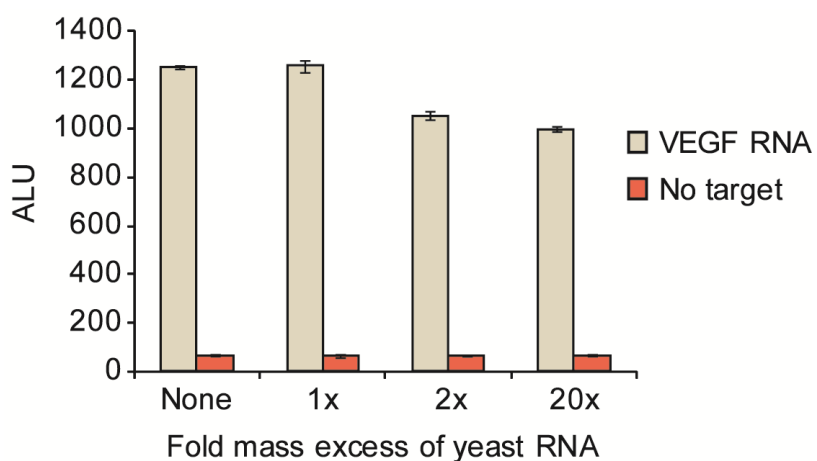


Figure S6. Effect of yeast RNA on VEGF annealing. VEGF was annealed to HP-guides in the presence of different mass excesses of yeast RNA. CLuciferase-E2C and Aart-NTLuciferase were reassembled in the presence of the annealed VEGF target at 1 nM (10 ng) with no yeast RNA or with 10 ng (1x), 20 ng (2x), or 200 ng (20x) yeast RNA, followed by luminescence readings. The data indicate a slight decrease in annealing at 20x yeast RNA, although higher amounts were not tested due to experimental constraints on higher yeast concentrations.

Complete Reference 1 Citations

(1) (a) Venter, J. C.; Adams, M. D.; Myers, E. W.; Li, P. W.; Mural, R. J.; Sutton, G. G.; Smith, H. O.; Yandell, M.; Evans, C. A.; Holt, R. A.; Gocayne, J. D.; Amanatides, P.; Ballew, R. M.; Huson, D. H.; Wortman, J. R.; Zhang, Q.; Kodira, C. D.; Zheng, X. Q. H.; Chen, L.; Skupski, M.; Subramanian, G.; Thomas, P. D.; Zhang, J. H.; Miklos, G. L. G.; Nelson, C.; Broder, S.; Clark, A. G.; Nadeau, C.; McKusick, V. A.; Zinder, N.; Levine, A. J.; Roberts, R. J.; Simon, M.; Slayman, C.; Hunkapiller, M.; Bolanos, R.; Delcher, A.; Dew, I.; Fasulo, D.; Flanigan, M.; Florea, L.; Halpern, A.; Hannenhalli, S.; Kravitz, S.; Levy, S.; Mobarry, C.; Reinert, K.; Remington, K.; Abu-Threideh, J.; Beasley, E.; Biddick, K.; Bonazzi, V.; Brandon, R.; Cargill, M.; Chandramouliswaran, I.; Charlab, R.; Chaturvedi, K.; Deng, Z. M.; Di Francesco, V.; Dunn, P.; Eilbeck, K.; Evangelista, C.; Gabrielian, A. E.; Gan, W.; Ge, W. M.; Gong, F. C.; Gu, Z. P.; Guan, P.; Heiman, T. J.; Higgins, M. E.; Ji, R. R.; Ke, Z. X.; Ketchum, K. A.; Lai, Z. W.; Lei, Y. D.; Li, Z. Y.; Li, J. Y.; Liang, Y.; Lin, X. Y.; Lu, F.; Merkulov, G. V.; Milshina, N.; Moore, H. M.; Naik, A. K.; Narayan, V. A.; Neelam, B.; Nusskern, D.; Rusch, D. B.; Salzberg, S.; Shao, W.; Shue, B. X.; Sun, J. T.; Wang, Z. Y.; Wang, A. H.; Wang, X.; Wang, J.; Wei, M. H.; Wides, R.; Xiao, C. L.; Yan, C. H. *Science* **2001**, *291*, 1304-1351.

(b) Sachidanandam, R.; Weissman, D.; Schmidt, S. C.; Kakol, J. M.; Stein, L. D.; Marth, G.; Sherry, S.; Mullikin, J. C.; Mortimore, B. J.; Willey, D. L.; Hunt, S. E.; Cole, C. G.; Coggill, P. C.; Rice, C. M.; Ning, Z. M.; Rogers, J.; Bentley, D. R.; Kwok, P. Y.; Mardis, E. R.; Yeh, R. T.; Schultz, B.; Cook, L.; Davenport, R.; Dante, M.; Fulton, L.; Hillier, L.; Waterston, R. H.; McPherson, J. D.; Gilman, B.; Schaffner, S.; Van Etten, W. J.; Reich, D.; Higgins, J.; Daly, M. J.; Blumenstiel, B.; Baldwin, J.; Stange-Thomann, N. S.; Zody, M. C.; Linton, L.; Lander, E. S.; Altshuler, D. *Nature* **2001**, *409*, 928-933.

Supplementary References

- (1) Meister, G.; Landthaler, M.; Patkaniowska, A.; Dorsett, Y.; Teng, G.; Tuschl, T. *Mol. Cell* **2004**, *15*, 185-197.
- (2) Ma, J.-B.; Ye, K.; Patel, D. J. *Nature* **2004**, *429*, 318-322.
- (3) Porter, J. R.; Stains, C. I.; Jester, B. W.; Ghosh, I. *J. Am. Chem. Soc.* **2008**, *130*, 6488-6497.
- (4) Segal, D. J.; Crotty, J. W.; Bhakta, M. S.; Barbas III, C. F.; Horton, N. C. *J. Mol. Biol.* **2006**, *363*, 405-421.
- (5) Beerli, R. R.; Segal, D. J.; Dreier, B.; Barbas III, C. F. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 14628-14633.
- (6) Cho, H.-S.; Mason, K.; Ramyar, K. X.; Stanley, A. M.; Gabelli, S. B.; Denney, D. W.; Leahy, D. J. *Nature* **2003**, *421*, 756-760.